Uniform Nonparametric Series Inference for Dependent Data with an Application to the Search and Matching Model*

Jia Li† and Zhipeng Liao‡

June 17, 2018

Abstract

This paper concerns the uniform inference for nonparametric series estimators in time-series applications. We develop a strong approximation theory of sample averages of serially dependent random vectors with dimensions growing with the sample size. The strong approximation is first proved for heterogeneous martingale difference arrays and then extended to general mixingales via martingale approximation, readily accommodating a majority of applications in applied econometrics. We use these results 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. The validity of high-dimensional heteroskedasticity and autocorrelation consistent (HAC) estimators is established for making feasible inference. The proposed method is broadly useful for forecast evaluation, empirical microstructure, dynamic stochastic equilibrium models and inference problems based on intersection bounds. We demonstrate the empirical relevance of the proposed method by studying the Mortensen–Pissarides search and matching model for equilibrium unemployment, and shed new light on the unemployment volatility puzzle from an econometric perspective.

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

---

*We thank Francesco Bianchi, Tim Bollerslev, Denis Chetverikov, Raffaella Giacomini, Cosmin Ilut, Kyle Jurado, Roger Koenker, Andrea Lanteri, Ulrich Mueller, Taisuke Otsu, Andrew Patton, Adam Rosen, George Tauchen and seminar participant at Duke, UCL and London School of Economics for their comments. Liao's research was partially supported by National Science Foundation Grant SES-1628889.

†Department of Economics, Duke University, Durham, NC 27708; e-mail: jl410@duke.edu.

‡Department of Economics, UCLA, Log Angeles, CA 90095; e-mail: zhipeng.liao@econ.ecla.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. 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), Chernozhukov, Chetverikov, and Kato (2014) 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, we develop a uniform inference theory for series estimators in time-series applications. Inspired by Chernozhukov, Lee, and Rosen (2013), our approach is also based on strong approximation, but with several distinct contributions unique to the time-series setting. Firstly, we prove a high-dimensional strong approximation (i.e., coupling) theorem in a general setting that accommodates typical time-series econometric applications. In this effort, we start with establishing a coupling result for general heterogeneous martingale difference arrays. Although this "baseline" result rules out serial correlation, it is actually useful in many applications in dynamic stochastic equilibrium (e.g., consumption-based asset pricing) models, in which an information flow is embedded. Going one step further, we use a martingale approximation technique to extend the martingale-difference coupling result to general mixingales, which cover most data generating processes in time-series econometrics as special cases, including martingale differences, ARMA processes, linear processes, various mixing and near-epoch dependent series. Equipped with these limit theorems, we establish a uniform inference theory for series estimators, which is our main econometric contribution. These results can be conveniently used for nonparametrically testing conditional moment equalities and, more generally, hypotheses based on intersection bounds (Chernozhukov, Lee, and Rosen (2013)). Finally, in order to conduct feasible inference, we prove the validity of classical Newey–West type HAC estimators for long-run covariance matrices with growing dimensions.

The proposed theory is broadly useful in many empirical time-series applications. In Section 2, we discuss how to use our method in a battery of prototype examples, along with heuristic discussions for the underlying econometric theory. This section provides practical guidance for the formal theory presented in Section 3. The applications are drawn from empirical microstructure,

[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]Yurinksii'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.

asset pricing and dynamic macroeconomic equilibrium models. As a concrete demonstration, we apply the proposed method to study the Mortensen–Pissarides search and matching model (Pissarides (1985), Mortensen and Pissarides (1994), Pissarides (2000)), which is the standard theory for equilibrium unemployment.

We focus empirically on the unemployment volatility puzzle. In an influential paper, Shimer (2005) showed 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 many references in Ljungqvist and Sargent (2017). Hagedorn and Manovskii (2008), on the other hand, took a different route and showed that the standard model actually can generate high levels of unemployment volatility using their alternative calibration strategy. The plausibility of this alternative calibration remains a contentious issue in the literature; see, for example, Hornstein, Krusell, and Violante (2005), Costain and Reiter (2008), Hall and Milgrom (2008) and Chodorow-Reich and Karabarbounis (2016).

To shed some light on this debate from an econometric perspective, we derive a conditional moment restriction from the equilibrium Bellman equations; using the proposed uniform inference method, we then test whether this restriction holds or not at the parameter values calibrated by Hagedorn and Manovskii (2008). The test strongly rejects the hypothesis that these calibrated values are compatible with the equilibrium conditional moment restriction. At the same time, we find a wide range of parameter values which the test does not reject, and use them to form an Anderson–Rubin confidence set. We further use this confidence set to impose an "admissibility" constraint on the parameter space. When the loss minimization in Hagedorn and Manovskii's calibration is constrained within this confidence set, the calibrated parameter values are notably different, both statistically and economically, from the unconstrained benchmark, leaving 26%–46% of unemployment volatility unexplained in the model. To this extent, the Shimer critique cannot be simply addressed by Hagedorn and Manovskii's alternative calibration once we impose the equilibrium conditional moment restriction. Our findings thus suggest that modifications to the standard Mortensen–Pissarides model are necessary for a better understanding of the cyclicality of unemployment.

The present paper is related to several strands of literature. The most closely related is the literature on series estimation and, more generally, sieve estimation. Early work in this area mainly focuses on semiparametric inference or pointwise nonparametric inference; see, for example, van de Geer (1990), Andrews (1991a), Gallant and Souza (1991), Newey (1997), Chen (2007), Chen and

Pouzo (2012), Chen, Liao, and Sun (2014), Chen and Pouzo (2015), Chen and Christensen (2015), Hansen (2015) and many references therein. Chernozhukov, Lee, and Rosen (2013), Chernozhukov, Chetverikov, and Kato (2014), Belloni, Chernozhukov, Chetverikov, and Kato (2015) and Chen and Christensen (2018) studied uniform inference for independent data. By contrast, our econometric focus is on the uniform series inference for time-series data.

On the technical side, our strong approximation results for heterogeneous martingale difference arrays and mixingales are related to the literature on high-dimensional coupling in statistics. The recent work of Chernozhukov, Lee, and Rosen (2013) and Belloni, Chernozhukov, Chetverikov, and Kato (2015) rely on Yurinskii's coupling (Yurinskii (1978)) for independent data. There has been limited research on high-dimensional coupling in the time-series setting. Zhang and Wu (2017) establish the strong approximation for the largest entry of a high-dimensional centered vector under a specific dependence structure based on stationary nonlinear systems (Wu (2005)); Chernozhukov, Chetverikov, and Kato (2013) show a similar result for mixing sequences. Unlike these papers, we consider the strong approximation for the entire high-dimensional vector for heterogeneous data with general forms of dependency (that are commonly used in time-series econometrics), establish the feasible uniform inference for the nonparametric series estimator and apply the theory to an important macroeconomic analysis.[3] Technically speaking, the martingale-based technique developed here is very different from the "large-block-small-block" technique employed in both Chernozhukov, Chetverikov, and Kato (2013) and Zhang and Wu (2017), and it is necessitated by the distinct dependence structure studied in the present paper. 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)).[4]

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. Zhang and Wu (2017) studied a "batched mean" estimator of high-dimensional long-run covariance matrices under a dependence structure based on nonlinear system theory. We focus on Newey–West

[3]The coupling for the largest entry of a high-dimensional vector can also be established in our setting by a straightforward adaptation of the theory developed here, which is available upon request.

[4]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.

type estimators and prove their validity under dependence structures that are commonly used in econometrics.

Finally, we contribute empirically to the search and matching literature, which is an important area in macroeconomics; see, for example, Pissarides (1985), Mortensen and Pissarides (1994), Shimer (2005), Hornstein, Krusell, and Violante (2005), Costain and Reiter (2008), Hagedorn and Manovskii (2008), Hall and Milgrom (2008), Pissarides (2009), Ljungqvist and Sargent (2017) and many references therein. Complementary to the standard calibration methodology that dominates quantitative work in this literature, we demonstrate how to use the proposed uniform inference method to help "disciplining" the calibration econometrically. Our approach of using Anderson–Rubin confidence sets for constraining the parameter space in the calibration is a new way of introducing econometric principles into an otherwise standard "computational experiment" (Kydland and Prescott (1996)). Our empirical findings shed light on the unemployment volatility puzzle from this new perspective. With this concrete demonstration, we hope to strengthen the message that modern econometric tools can be fruitfully used to assist quantitative analysis of dynamic stochastic macroeconomic equilibrium models.

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 formal econometric theory. The empirical application on equilibrium unemployment is given in Section 4. Section 5 concludes. Technical derivations, including 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 jth 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, $\|X\|_p = (\mathbb{E}\,\|X\|^p)^{1/p}$.

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 $\widetilde{S}_n$ of random vectors, defined on the same probability space, is called a *strong approximation* of S_n if

$$\|S_n - \widetilde{S}_n\| = o_p(\delta_n) \tag{2.1}$$

for some real sequence $\delta_n \to 0$; we reserve the symbol δ_n for this role throughout.[5] A useful special case is when the approximating variable $\widetilde{S}_n$ has a Gaussian $\mathcal{N}(0, \Sigma_n)$ distribution with some $m_n \times m_n$ covariance matrix Σ_n, so that (2.1) formalizes a notion of "asymptotic normality" for the random vector S_n; we refer to Σ_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."[6]

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).

[5]We note that without specifying δ_n, condition (2.1) is equivalent to $\|S_n - \widetilde{S}_n\| = o_p(1)$. 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 δ_n is needed explicitly for justifying feasible inference (by relying on anit-concentration inequalities) in the high-dimensional case.

[6]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 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.

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), 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; see Andrews (1991a) and Belloni, Chernozhukov, Chetverikov, and Kato (2015) for many references. 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:

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

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 $[\widehat{L}_n(x), \widehat{U}_n(x)]$ such that the uniform coverage probability

$$\mathbb{P}\left(\widehat{L}_n(x) \leq h\left(x\right) \leq \widehat{U}_n(x) \text{ for all } x \in \mathcal{X}\right) \tag{2.3}$$

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.[7] 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),

[7] 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.

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 $\hat{h}_n(x) \equiv P(x)^\top \hat{b}_n$, where $\hat{b}_n$ is the least-square coefficient obtained from regressing Y_t on $P(X_t)$, that is,

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

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 $h(x) - P(x)^\top b_n^*$ is close to zero uniformly in x as $m_n \to \infty$; such an approximation is justified by numerical approximation theory.[8] The sampling error of $\hat{b}_n$ is measured by

$$S_n = n^{1/2}(\hat{b}_n - b_n^*).$$

Based on the strong approximation for the sample average of $P(X_t)u_t$, we can construct a strong Gaussian approximation $\widetilde{S}_n$ for S_n such that $\widetilde{S}_n \sim \mathcal{N}(0, \Sigma_n)$. 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 function $\sigma_n(\cdot)$ can be estimated "sufficiently well" by a sample-analogue estimator $\hat{\sigma}_n(\cdot)$, which generally involves a high-dimensional HAC estimator (see Section 3.3).

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

$$\frac{n^{1/2}\left(\hat{h}_n(x) - h(x)\right)}{\hat{\sigma}_n(x)} = \frac{P(x)^\top \widetilde{S}_n}{\sigma_n(x)} + O_p(\delta_n), \tag{2.5}$$

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 $\widetilde{S}_n$ but with different loadings (i.e., $P(x)/\sigma_n(x)$). Importantly, (2.5) depicts the asymptotic

[8] We remind the reader that the "true" parameter b_n^* depends on n because its dimension (i.e., m_n) grows asymptotically; see Assumption 6(i) for the formal definition of b_n^*.

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 3.4.

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.

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

$$\mathbb{E}\left[g\left(Y_t^*, \gamma_0\right)|X_t\right] = 0, \tag{2.6}$$

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 γ_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 γ_0, and will return to the case with unknown γ_0 at the end of this subsection.

To implement the test, we cast (2.6) as a nonparametric regression in the form of (2.2) by setting $Y_t = g(Y_t^*, \gamma_0)$, $h(x) = \mathbb{E}[Y_t|X_t = x]$ and $u_t = Y_t - h(X_t)$. Testing the conditional moment restriction (2.6) 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,

$$\widehat{L}_n(x) \leq 0 \leq \widehat{U}_n(x) \text{ for all } x \in \mathcal{X}. \tag{2.7}$$

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.

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 $Y_t^* = (C_t, C_{t+1}, R_{t+1})$)

$$g(Y_t^*, \gamma_0) = \frac{\delta u'(C_{t+1})}{u'(C_t)} R_{t+1},$$

where R_{t+1} is the excess return of an asset, C_t is the consumption, δ is the discount rate and $u'(\cdot)$ is the marginal utility function parameterized by γ. 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 (see Section 4) on the search and matching model for equilibrium unemployment is of this type. To avoid repetition, we refer the reader to Section 4 for details.

Finally, we return to the issue with unknown γ_0. In this case, γ_0 should be replaced by an estimated or, perhaps more commonly in macroeconomic applications, a calibrated value $\hat{\gamma}_n$.[9] 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 3.5, the estimation/calibration error in $\hat{\gamma}_n$ is asymptotically negligible under empirically plausible conditions. The intuition is straightforward: since γ_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 γ_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

[9]We refer the reader to the comprehensive review of Dawkins, Srinivasan, and Whalley (2001) for discussions about estimation and calibration.

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 Main theoretical results

This section contains our theoretical results. Section 3.1 and Section 3.2 present the strong approximation theorems for heterogeneous martingale differences and mixingales, respectively. Section 3.3 establishes the validity of classical HAC estimators in the high-dimensional setting. The uniform inference theory for series estimators is presented in Section 3.4. Section 3.5 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 heterogeneity. Although the martingale difference is uncorrelated, it can accommodate general forms of dependence through high-order conditional moments. This result will be extended to mixingales in Section 3.2, below.[10]

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 $\mathbb{E}\left[X_{n,t}|\mathcal{F}_{n,t-1}\right] = 0$. Let $V_{n,t} \equiv \mathbb{E}\left[X_{n,t}X_{n,t}^\top|\mathcal{F}_{n,t-1}\right]$ denote the conditional covariance matrix of $X_{n,t}$ and set

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

In typical applications, $k_n = n$ is the sample size and $X_{n,t} = k_n^{-1/2}X_t$ represents a normalized version of the series X_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.

[10] More generally, the martingale-difference coupling developed here may be extended to the setting with discretized semimartingales (see, e.g., Jacod and Protter (2012) and Aït-Sahalia and Jacod (2014)) that are routinely used in high-frequency econometrics.

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

$$S_n \xrightarrow{d} \mathcal{N}(0, \Sigma), \tag{3.1}$$

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 $\widetilde{S}_n \sim \mathcal{N}(0, \Sigma_n)$ such that $\|S_n - \widetilde{S}_n\| = O_p(\delta_n)$ 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 $h_n/k_n \to 1$,*

$$\left\| \sum_{t=1}^{h_n} V_{n,t} - \Sigma_{n,h_n} \right\|_S = O_p(r_n). \tag{3.2}$$

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) is somewhat non-standard and needs further discussion. When $h_n = k_n$, (3.2) requires the conditional covariance of the martingale S_n (i.e., $\sum_{t=1}^{k_n} V_{n,t}$) to be close to the pre-asymptotic covariance matrix Σ_n. This condition would be easily verified by appealing to a law of large numbers in conventional settings with fixed dimensions, but this argument needs to be adapted slightly to accommodate the growing dimension in the present setting. More generally, we require (3.2) holds for any h_n that is bounded by and is close to k_n. This requirement typically does not complicate the verification of condition (3.2), but it is needed in our proof which relies on a stopping time technique on large matrices for constructing the coupling variable $\widetilde{S}_n$. This technical complication arises because the conditional covariance $V_{n,t}$ is generally stochastic in the time-series setting, whereas it would be nonrandom for independent data.

Assumption 1 is easy to verify under primitive conditions, with condition (ii) being the relatively nontrivial part. For concreteness, we illustrate how to verify condition (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 1. *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 $\sup_{t,j,l} \|v_t^{(j,l)}\|_q \leq \bar{c}_n^2$ 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 α_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 ϕ_k satisfying $\sum_{k=1}^{k_n} \phi_k^{1/2} < \infty$. Then, uniformly for all sequence h_n that satisfies $h_n \leq k_n$,*

$$\left\| \sum_{t=1}^{h_n} \left(V_{n,t} - \mathbb{E}\left[V_{n,t}\right]\right) \right\|_2 = O(r_n), \quad \textit{for} \quad r_n \equiv \bar{c}_n^2 m_n k_n^{-1/2}. \tag{3.3}$$

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

Comment. The sequence $\bar{c}_n$ 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 $X_{n,t} = u_t P(X_t) k_n^{-1/2}$. We have $\bar{c}_n = O(1)$ if $P(\cdot)$ collects splines or trigonometric polynomials and $\bar{c}_n = O(m_n^{1/2})$ if $P(\cdot)$ consists of power series or Legendre polynomials. In these two cases, $r_n = o(1)$ is implied by $m_n \ll k_n^{1/2}$ and $m_n \ll k_n^{1/4}$, respectively.

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

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

$$\left\| S_n - \widetilde{S}_n \right\| = O_p(m_n^{1/2} r_n^{1/2} + (B_n m_n)^{1/3}), \tag{3.4}$$

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

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* Σ_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. 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 Σ_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. 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 Σ_n (which is nonrandom), we can use a version of Lindeberg's method and Strassens' theorem for establishing the strong approximation.

The strong approximation rate in (3.4) can be simplified under additional (mild) assumptions. Corollary 1, 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 1. *Under the same setting as Theorem 1, if* $\sup_{t,j} \mathbb{E}[(X_{n,t}^{(j)})^4] = O\left(k_n^{-2}\right)$ *holds in addition, then* $B_n = O(k_n^{-1/2} m_n^{3/2})$. *Consequently,* $\|S_n - \widetilde{S}_n\| = O_p(m_n^{1/2} r_n^{1/2} + m_n^{5/6} k_n^{-1/6})$.

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.

3.2 Strong approximation for mixingales via high-dimensional martingale approximation

Theorem 1 is apparently restrictive for time-series applications since martingale differences are serially uncorrelated. In this subsection, we extend the coupling result above towards mixingale processes via martingale approximation. Mixingales form a quite 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; see, for example, Davidson (1994) for a comprehensive review.[11] The coupling result developed here thus readily accommodates most applications in macroeconomics and finance.

We now turn to the formal setup. 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 $k \geq 0$,

$$\left\| \mathbb{E}[X_{n,t}^{(j)} | \mathcal{F}_{n,t-k}] \right\|_q \leq c_{n,t} \psi_k, \quad \left\| X_{n,t}^{(j)} - \mathbb{E}[X_{n,t}^{(j)} | \mathcal{F}_{n,t+k}] \right\|_q \leq c_{n,t} \psi_{k+1}, \tag{3.5}$$

where the constants $c_{n,t}$ and ψ_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 $k_n^{1/2} X_{n,t}$ (recall the comment following Proposition 1).

Assumption 2. *The array* $(X_{n,t})$ *satisfies (3.5) 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 2 allows us to approximate the partial sum of the mixingale $X_{n,t}$ using a martingale. More precisely, we can represent

$$X_{n,t} = X_{n,t}^* + \widetilde{X}_{n,t} - \widetilde{X}_{n,t+1} \tag{3.6}$$

[11] It is well known that linear processes and mixing processes are special cases of mixingales. Under certain conditions, near-epoch dependent arrays also form mixingales; see, for example, Theorem 17.5 of Davidson (1994). More generally, it may be possible to extend the martingale-difference coupling result to an even larger class of data generating processes than the mixingale class, provided that a martingale approximation result is available.

where $X_{n,t}^* \equiv \sum_{s=-\infty}^{\infty} \{\mathbb{E}[X_{n,t+s}|\mathcal{F}_{n,t}] - \mathbb{E}[X_{n,t+s}|\mathcal{F}_{n,t-1}]\}$ forms a martingale difference and the "residual" variable $\widetilde{X}_{n,t}$ satisfies $\sup_{j,t} \|\widetilde{X}_{n,t}^{(j)}\|_2 = O(\bar{c}_n k_n^{-1/2})$.[12] This representation further permits an approximation of S_n via the martingale $S_n^* = \sum_{t=1}^{k_n} X_{n,t}^*$, that is,

$$\|S_n - S_n^*\|_2 = \left\|\widetilde{X}_{n,1} - \widetilde{X}_{n,k_n+1}\right\|_2 = O(\bar{c}_n m_n^{1/2} k_n^{-1/2}). \tag{3.7}$$

In the typical case with $\bar{c}_n = O(1)$, the approximation error in (3.7) 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 2 formalizes this result under a high-level condition (see condition (ii) below) regarding the approximating martigale difference array $X_{n,t}^*$. In Supplemental Appendix S.B.1, we illustrate how to verify this high-level condition with concrete examples.

Theorem 2. *Suppose (i) Assumption 2 holds; (ii) Assumption 1 is satisfied for the martingale difference array $X_{n,t}^*$; and (iii) the largest eigenvalue of Σ_n is bounded. Then there exists a sequence $\widetilde{S}_n$ of m_n-dimensional random vectors with distribution $\mathcal{N}(0, \Sigma_n)$ such that*

$$\left\|S_n - \widetilde{S}_n\right\| = O_p(\bar{c}_n m_n^{1/2} k_n^{-1/2}) + O_p(m_n^{1/2} r_n^{1/2} + (B_n^* m_n)^{1/3}) + O_p(\bar{c}_n m_n k_n^{-1/2} + \bar{c}_n^2 m_n^{3/2} k_n^{-1}), \tag{3.8}$$

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

Comments. (i) There are three types of approximation errors underlying this strong approximation result. The first $O_p(\bar{c}_n m_n^{1/2} k_n^{-1/2})$ component is due to the martingale approximation. The second term arises from the approximation of the martingale S_n^* using a centered Gaussian variable $\widetilde{S}_n^*$ with covariance matrix $\Sigma_n^* \equiv \mathbb{E}[S_n^* S_n^{*\top}]$. The magnitude of this error is characterized by Theorem 1 as $O_p(m_n^{1/2} r_n^{1/2} + (B_n^* m_n)^{1/3})$. The third error component measures the distance between the two coupling variables $\widetilde{S}_n^*$ and $\widetilde{S}_n$, and is of order $O_p(\bar{c}_n m_n k_n^{-1/2} + \bar{c}_n^2 m_n^{3/2} k_n^{-1})$.

(ii) It is instructive to simplify the rate in (3.8) in a "typical" setting with $\bar{c}_n = O(1)$ and $m_n k_n^{-1} = O(1)$. Corollary 1 suggests that $B_n^* = O(k_n^{-1/2} m_n^{3/2})$. We then deduce

$$\left\|S_n - \widetilde{S}_n\right\| = O_p(m_n^{1/2} r_n^{1/2} + m_n^{5/6} k_n^{-1/6}) + O_p(m_n k_n^{-1/2}). \tag{3.9}$$

Since the validity of the strong approximation requires $m_n \ll k_n^{1/5}$, $m_n k_n^{-1/2} = o(m_n^{5/6} k_n^{-1/6})$. We can thus simplify the error bound as $\|S_n - \widetilde{S}_n\| = O_p(m_n^{1/2} r_n^{1/2} + m_n^{5/6} k_n^{-1/6})$, which coincides with the rate shown in Theorem 1. In this sense, our generalization of the strong approximation result from martingale differences towards mixingales typically leads to no additional cost in terms of convergence rate.

[12] See Lemma A3 in the supplemental appendix for technical details about this approximation.

The strong approximation results established in Theorems 1 and 2 are mainly used in the current paper for establishing a uniform inference theory for series estimators. That noted, these results are also useful in other econometric settings. One case in point is the "reality check" of White (2000) for testing superior performance of competing models; also see Romano and Wolf (2005), Hansen (2005) and Hansen, Lunde, and Nason (2011) for refinements and extensions. The asymptotic properties of such tests rely on the asymptotic normality of the sample average of a loss differential vector that summarizes the relative performance of competing models. While the existing theory in aforementioned work has been developed for a fixed number of models, practical applications often involve "many" models. Indeed, White (2000) suggested (p. 1111) that this feature could be captured by letting the number of models (i.e., m_n) grow with the sample size. The strong approximation result developed here may be used to address the technical complication due to the growing dimension. But we do not pursue such results formally so as to remain focused on series inference.

3.3 High-dimensional HAC estimation

In this subsection, we establish the asymptotic validity of a class of HAC estimators for the covariance matrix Σ_n which is needed for conducting feasible inference. 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. 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 5(b)).

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 $\widetilde{\Gamma}_{X,n}(s)$, as

$$\widetilde{\Gamma}_{X,n}(s) \equiv \sum_{t=1}^{k_n - s} X_{n,t} X_{n,t+s}^{\top} \tag{3.10}$$

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

$$\widetilde{\Sigma}_n \equiv \sum_{s=-k_n+1}^{k_n-1} \mathcal{K}(s/M_n)\, \widetilde{\Gamma}_{X,n}(s) \tag{3.11}$$

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 3. *(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 $r_1 \in (0, \infty]$, $\lim_{x \to 0}(1 - \mathcal{K}(x))/|x|^{r_1} = C$.*[13]

In order to analyze the limit behavior of $\widetilde{\Gamma}_{X,n}(s)$ 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 (3.5)). More precisely, we maintain the following assumption that can be easily verified under more primitive conditions.

Assumption 4. *We have Assumption 2. Moreover, (i) for any $n > 0$, any t and any j, $\mathbb{E}[X_{n,t}] = 0$ and $\mathbb{E}[X_{n,t}X_{n,t+j}]$ only depends on n and j; (ii) for all $j \geq 0$ and $s \geq 0$,*

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

(iii) $\sup_t \max_{1 \leq l,k \leq m_n} \left\| X_{n,t}^{(k)} X_{n,t+j}^{(l)} \right\|_2 \leq \bar{c}_n^2 k_n^{-1}$ for all $j \geq 0$; (iv) $\sup_{s \geq 0} s\psi_s^2 < \infty$ and $\sum_{s=0}^{\infty} s^{r_2}\psi_s < \infty$ 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 $X_{n,t}X_{n,t+j}^{\top}$.[14] Conditions (iii) reflects that the scale of $k_n^{1/2}X_{n,t}$ is bounded by $\bar{c}_n$. Condition (iv) specifies the level of weak dependence. The rate of convergence of the HAC estimator is given by the following theorem.

Theorem 3. *Under Assumptions 3 and 4, $\|\widetilde{\Sigma}_n - \Sigma_n\| = O_p(\bar{c}_n^2 m_n (M_n k_n^{-1/2} + M_n^{-r_1 \wedge r_2}))$.*

COMMENT. Theorem 3 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 $\bar{c}_n = O(1)$, the convergence rate is simply $O_p(M_n k_n^{-1/2} + M_n^{-r_1 \wedge r_2})$. In this special case, $\widetilde{\Sigma}_n$ 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 m_n.

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

[13] 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.

[14] 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 ψ_s so as to simplify the technical exposition.

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 3 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 θ. The unknown parameter θ_0 may be finite or infinite dimensional and can be estimated by $\hat{\theta}_n$. We use $\widehat{X}_{n,t} = k_n^{-1/2} g(Z_t, \hat{\theta}_n)$ as a proxy for $X_{n,t}$. The feasible versions of (3.10) and (3.11) are then given by

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

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

Theorem 4, below, characterizes the convergence rate of the feasible HAC estimator $\widehat{\Sigma}_n$ when $\hat{\theta}_n$ is "sufficiently close" to the true value θ_0; the latter condition is formalized as follows.

Assumption 5. *(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 5(i) is a high-level condition that embodies two types of regularities: the smoothness of $g(\cdot)$ with respect to θ and the convergence rate of the preliminary estimator $\hat{\theta}_n$. Quite commonly, $g(\cdot)$ is stochastically Lipschitz in θ and $\delta_{\theta,n}$ equals the convergence rate of $\hat{\theta}_n$. Sharper primitive conditions might be tailored in more specific applications. Assumption 5(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 4. *Under Assumptions 3, 4 and 5, we have*

$$\left\| \widehat{\Sigma}_n - \Sigma_n \right\| = O_p\left(\bar{c}_n^2 m_n (M_n k_n^{-1/2} + M_n^{-r_1 \wedge r_2}) \right) + O_p\left(M_n m_n^{1/2} \delta_{\theta,n} \right). \tag{3.12}$$

Comments. (i) The estimation error shown in (3.12) contains two components. The first term accounts for the estimation error in the infeasible estimator $\widetilde{\Sigma}_n$ 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 (3.12) can be further simplified when θ 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

(3.12) 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 θ_0. When θ 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.

3.4 Uniform inference for nonparametric series regressions

In this subsection, we apply the limit theorems 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 $1 \leq t \leq n$,

$$Y_t = h\left(X_t\right) + u_t \tag{3.13}$$

where $h\left(\cdot\right)$ 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

$$\mathbb{E}\left[u_t | X_t\right] = 0. \tag{3.14}$$

Dynamic stochastic equilibrium models often imply a stronger restriction

$$\mathbb{E}\left[u_t | \mathcal{F}_{t-1}\right] = 0, \tag{3.15}$$

where the information flow $\mathcal{F}_{t-1}$ is a σ-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 $h\left(x\right)$ is given by $\hat{h}_n(x) \equiv P\left(x\right)^\top \hat{b}_n$, where $P\left(\cdot\right)$ collects the basis functions and $\hat{b}_n$ is the least-square coefficient obtained by regressing Y_t on $P\left(X_t\right)$; 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 $\Sigma_n \equiv Q_n^{-1} A_n Q_n^{-1}$, where

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

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

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

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

$$\widehat{Q}_n \equiv n^{-1}\sum_{t=1}^{n}P(X_t)P(X_t)^\top.$$

For the estimation of A_n, we consider two scenarios. The first scenario is when u_t forms a martingale difference sequence, that is, (3.15) holds. In this case, $A_n = n^{-1}\sum_{t=1}^{n}\mathbb{E}\left[u_t^2 P(X_t)P(X_t)^\top\right]$ and it can be estimated by

$$\widehat{A}_n \equiv n^{-1}\sum_{t=1}^{n}\widehat{u}_t^2 P(X_t)P(X_t)^\top, \quad \text{where} \quad \widehat{u}_t = Y_t - \hat{h}_n(X_t). \tag{3.16}$$

In the second scenario, we suppose only the mean independence assumption (3.14), so A_n is generally a long-run covariance matrix. We use a HAC estimator for A_n as described in Section 3.3. We set

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

and estimate A_n using

$$\widehat{A}_n \equiv n^{-1}\sum_{s=-n+1}^{n-1}\mathcal{K}\left(s/M_n\right)\widehat{\Gamma}_n(s). \tag{3.17}$$

With $\widehat{\Sigma}_n \equiv \widehat{Q}_n^{-1}\widehat{A}_n\widehat{Q}_n^{-1}$, the estimator of $\sigma_n\left(x\right)$ is given by

$$\hat{\sigma}_n\left(x\right) \equiv \left(P\left(x\right)^\top\widehat{\Sigma}_n P\left(x\right)\right)^{1/2}.$$

Under some regularity conditions, we shall show (see Theorem 5) that the "sup-t" statistic

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

can be (strongly) approximated by

$$\widetilde{T}_n \equiv \sup_{x\in\mathcal{X}}\left|\frac{P(x)^\top\widetilde{S}_n}{\sigma_n\left(x\right)}\right|, \quad \widetilde{S}_n \sim \mathcal{N}\left(0,\Sigma_n\right).$$

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

Algorithm 1 (Uniform confidence band construction)

Step 1. Draw m_n-dimensional standard normal vectors ξ_n repeatedly and compute

$$\widetilde{T}_n^* \equiv \sup_{x\in\mathcal{X}}\left|\frac{P(x)^\top\widehat{\Sigma}_n^{1/2}\xi_n}{\hat{\sigma}_n\left(x\right)}\right|.$$

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

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

We are now ready to present the asymptotic theory that justifies the validity of the confidence band described in the algorithm above. To streamline the discussion, 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 that we have developed in the previous subsections. Below, we denote $\zeta_n^L \equiv \sup_{x_1,x_2\in\mathcal{X}} \|P(x_1) - P(x_2)\| / \|x_1 - x_2\|$.

Assumption 6. *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*

$$\sup_{x\in\mathcal{X}} \left(1 + \|P(x)\|^{-1}\right) n^{1/2} \left|h(x) - P(x)^\top b_n^*\right| = O(\delta_{1,n});$$

(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 $\widetilde{N}_n \sim \mathcal{N}(0, A_n)$ such that

$$\left\| n^{-1/2} \sum_{t=1}^n P(X_t)u_t - \widetilde{N}_n \right\| = O_p(\delta_{2,n});$$

(iv) $\|\widehat{Q}_n - Q_n\| = O_p(\delta_{3,n})$; (v) $\|\widehat{A}_n - A_n\| = O_p(\delta_{4,n})$.

A few remarks on Assumption 6 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. The strong approximation in condition (iii) can be verified by using Theorem 2 in general and, if u_t is a martingale difference sequence (i.e., (3.15) holds), it suffices to apply Theorem 1 to the martingale difference array $X_{n,t} = n^{-1/2}P(X_t)u_t$. Conditions (iv) and (v) pertain to the convergence rates of $\widehat{Q}_n$ and $\widehat{A}_n$. Theorem 4 can be used to provide the rate for $\widehat{A}_n$. The convergence rate for $\widehat{Q}_n$ can be derived in a similar (actually simpler) fashion.[15]

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

Theorem 5. *The following statements hold under Assumption 6:*

[15]Primitive conditions for Assumption 6 are provided and justified in Supplemental Appendix S.B.2.

(a) the sup-t statistic $\widehat{T}_n$ admits a strong approximation, that is, $\widehat{T}_n = \widetilde{T}_n + O_p(\delta_n)$ 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 $1-\alpha$:

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

3.5 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

$$\mathbb{E}\left[g(Y_t^*, \gamma_0)|X_t\right] = 0, \tag{3.18}$$

where $g(\cdot)$ is a known function and γ_0 is a finite-dimensional parameter from a parameter space $\Upsilon \subseteq \mathbb{R}^d$. As discussed in Section 2.3, when γ_0 is known, we can cast the testing problem as a nonparametric regression by setting

$$Y_t = g(Y_t^*, \gamma_0),\ h(x) = \mathbb{E}\left[Y_t|X_t = x\right] \text{ and } u_t = Y_t - \mathbb{E}\left[Y_t|X_t\right]. \tag{3.19}$$

The test for (3.18) can then be carried out by examining whether the uniform confidence bound $[\widehat{L}_n(x), \widehat{U}_n(x)]$ covers the zero function (recall (2.7)).

This testing strategy is inspired by Chernozhukov, Lee, and Rosen (2013). These authors conduct inference for intersection bounds using the uniform inference theory for series estimators, including the inference for conditional moment inequalities as a special case. Although we restrict attention to conditional moment equalities, our technical results for the strong approximation of the t-statistic process (i.e., $n^{1/2}(\hat{h}_n(\cdot) - h(\cdot))/\hat{\sigma}_n(\cdot)$) and the standard error estimator $\hat{\sigma}_n(\cdot)$ can actually be used to verify the key high-level conditions in Chernozhukov, Lee, and Rosen (2013) and, hence, to extend their method to time-series applications (see Supplemental Appendix S.B.3 for the formal result).[16] Like Chernozhukov, Lee, and Rosen (2013), our nonparametric

[16] The "main and preferred approach (p. 690)" of Chernozhukov, Lee, and Rosen (2013) is given by their Theorem 2, which relies on Conditions C.1–C.4 in that paper. These authors show (Lemma 5) that these high-level conditions are implied by Condition NS in the context of series estimation. To extend their result to the time-series setting, we only need to verify Condition NS(i)(a) and NS(ii) in that paper. The former concerns the strong approximation of the t-statistic process and the latter is on the convergence rate of the covariance matrix estimator. Both conditions can be verified under Assumption 7 below (see Proposition B3 in the supplemental appendix of the current paper for technical details).

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 instruments. These two approaches are complementary with their own merits; see Chernozhukov, Lee, and Rosen (2013) for further discussions.

The situation becomes somewhat more complicated when γ_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 γ_0 is identified from the conditional moment restriction (3.18) 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.[17]

Equipped with $\hat{\gamma}_n$, we can implement the econometric procedure described in Section 3.4, except that we take Y_t as the "generated" variable $g(Y_t^*, \hat{\gamma}_n)$. More precisely, we set

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

$\hat{h}_n(x) = P(x)^\top \hat{b}_n$, $\widehat{u}_t = g(Y_t^*, \hat{\gamma}_n) - \hat{h}_n(X_t)$ and then define $\hat{\sigma}_n(x)$ similarly as in Section 3.4. As alluded to previously (see Section 2.3), we aim to provide sufficient conditions such that replacing γ_0 with $\hat{\gamma}_n$ leads to negligible errors. The intuition is that the parametric proxy error in $\hat{\gamma}_n$ tends to be asymptotically dominated by the "statistical noise" in the nonparametric test.[18] We formalize this intuition with a few assumptions.

Assumption 7. *Conditions (i)-(iv) of Assumption 6 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 6 holds for $\widehat{A}_n$ defined using $\widehat{u}_t = g(Y_t^*, \hat{\gamma}_n) - \hat{h}_n(X_t)$, and $\delta_n(\log m_n)^{1/2} = o(1)$.*

Assumption 7 allows us to cast the testing problem into the nonparametric regression setting of Section 3.4. 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

[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.

[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.

approximation of the infeasible estimator defined using $g\left(Y_t^*, \gamma_0\right)$ 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 $g_\gamma\left(\cdot\right)$ and $g_{\gamma\gamma}\left(\cdot\right)$ to denote the first and the second partial derivatives of $g\left(y, \gamma\right)$ with respect to γ, and we set

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

Assumption 8. *Suppose (i) for any y, $g(y,\gamma)$ is twice continuously differentiable with respect to γ; (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 ϕ_n^*, $\sup_{x\in\mathcal{X}} \left\|P(x)^\top \phi_n^* - H(x)\right\| = O(m_n^{-\rho})$; (iv) $\sup_{\gamma\in\Upsilon} n^{-1} \sum_{t=1}^{n} \left\|g_{\gamma\gamma}(Y_t^*, \gamma)\right\|^2 = O_p(1)$, $\sup_{x\in\mathcal{X}} \|H(x)\| < \infty$ and $\mathbb{E}[\|g_\gamma(Y_t^*, \gamma_0)\|^2]$ is bounded; (v) $\max_{1\le k\le m_n} \sup_{x\in\mathcal{X}} |p_k(x)| \le \zeta_n$ for a non-decreasing positive sequence $\zeta_n = O(m_n^{\rho-1/2})$; (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 8 jointly impose a type of (stochastic) smoothness for the moment functions with respect to γ. These conditions are useful for controlling the effect of 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 γ_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.

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 $\widehat{T}_n$ and the critical value $cv_{n,\alpha}$.

Step 2. Reject the null hypothesis (3.18) at significance level α if $\widehat{T}_n > cv_{n,\alpha}$. □

Theorem 6. *Suppose that Assumptions 7 and 8 hold. Then under the null hypothesis (3.18), the test described in Algorithm 2 has asymptotic level α. Under the alternative hypothesis that $\mathbb{E}\left[g(Y_t^*, \gamma_0)|X_t = x\right] \neq 0$ for some $x \in \mathcal{X}$, the test rejects with probability approaching one.*

4 Empirical application on a search and matching model

4.1 The model and the equilibrium conditional moment restriction

The Mortensen–Pissarides search and matching model (Pissarides (1985), Mortensen and Pissarides (1994), Pissarides (2000)) has become the standard theory of equilibrium unemployment. 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. For example, Shimer (2004) and Hall (2005) first introduce sticky wages; Hall and Milgrom (2008) replace the standard Nash model of bargaining with an alternating offer bargaining model and Mortensen and Nagypál (2007) further consider separation shocks in this setting; Gertler and Trigari (2009) model sticky wage via staggered multiperiod wage contracting. Pissarides (2009) emphasizes the distinction between wage stickiness in continuing jobs and that in new matches and instead proposes a model with fixed matching cost. Ljungqvist and Sargent (2017) provide additional references and identify the common channel of these reconfigured matching models using the concept of fundamental surplus fraction.

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 that in turn greatly improves the standard model's capacity for generating unemployment volatility. To the extent that this alternative calibration is plausible, the Shimer critique to the standard model is less of a concern.

Several authors have argued that HM's calibrated value of the nonmarket activity is implausibly large according to certain calibration metrics. For example, Hall and Milgrom (2008) state that HM's calibrated value of the 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 an effect of the unemployment insurance policy much higher than their empirical estimates. These arguments are of course economically sound in principle, but 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.[19] To appreciate how complex the calibration of this single parameter can be, we cite the recent comprehensive study of Chodorow-Reich and Karabarbounis (2016). To measure the nonmarket return, these authors resort to a broad range of data sources, including the Current Population Survey, the Survey of Income and Program Participation, Interval Revenue Service Public Use Files, the Consumer Expenditure Survey, the Panel Study of Income Dynamics, and National Income and Product Accounts data. Chodorow-Reich and Karabarbounis (2016) find that, depending on the specific auxiliary assumptions on the utility function, the value of nonmarket activity can range from 0.47 to 0.96; in particular, this range contains HM's calibrated value of 0.955. How to calibrate the value of this parameter (among many other parameters) still appears to be a contentious issue in the literature; see Hornstein, Krusell, and Violante (2005) for a review.

We aim to shed some light on this debate. Rather than resorting to some "external" target like in aforementioned calibration exercises, we rely on a conditional moment restriction that arises "internally" from the equilibrium Bellman equations.[20] We then test it using the proposed nonparametric test as described in Subsection 3.5. Constructively, by forming an Anderson–Rubin confidence set, we explore the extent to which this equilibrium conditional moment restriction restricts a calibrated model's capacity for generating unemployment volatility. We stress from the outset that, our approach should be considered as a *complement*, instead of a substitute, to the conventional paradigm for evaluating calibrated models. Given the complexity of the empirical problem of interest, we believe that the new econometric perspective offered here is a useful addition to empirical macroeconomists' toolbox.

We now turn to the details. For the purpose at hand, we 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

[19] More precisely, Hall and Milgrom (2008) rely on a specific utility function calibrated using additional data. In a comprehensive study, Chodorow-Reich and Karabarbounis (2016) show that the calibration of the nonmarket return is sensitive to the choice of utility function; see Section IV.E in that paper. Costain and Reiter (2008) rely on econometric estimates of the effect of unemployment insurance benefit. This estimation problem itself is empirically difficult and controversial as explained in Section 4 of Costain and Reiter (2008); see also the response of Hagedorn and Manovskii (2008), p. 1703.

[20] We use the word "internal" in two literal senses. First, our test can be performed without making additional economic assumptions (e.g., the form of utility function as needed in Hall and Milgrom (2008)). Second, we do not use additional data (e.g., the impact of unemployment insurance across countries as used by Costain and Reiter (2008)). In fact, we only use data on productivity and labor market tightness, which are shared among all quantitative research in this area.

the same discount factor δ. 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:

$$\log p_{t+1} = \rho \log p_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right).$$

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, where the subscript p indicates that c_p is a function of productivity. HM parameterize the vacancy cost c_p as

$$c_p \equiv c^K p + c^W p^\xi, \tag{4.1}$$

where c^K and c^W are the steady-state capital cost and labor cost for posting vacancies, respectively, and ξ is the wage-productivity elasticity.

Let u_t denote the unemployment rate and v_t the number of vacancies. The number of new matches is given by the matching function

$$m\left(u_t, v_t\right) = \frac{u_t v_t}{\left(u_t^l + v_t^l\right)^{1/l}},$$

for some matching parameter $l > 0$; den Haan, Ramey, and Watson (2000) provide motivations for using this matching function (see their footnote 6). 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,

$$f\left(\theta_t\right) \equiv \frac{m\left(u_t, v_t\right)}{u_t} = \frac{\theta_t}{\left(1 + \theta_t^l\right)^{1/l}}, \quad q\left(\theta_t\right) \equiv \frac{m\left(u_t, v_t\right)}{v_t} = \frac{1}{\left(1 + \theta_t^l\right)^{1/l}}.$$

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 $\mathbb{E}_p[X_{p'}]$ denote the one-period ahead conditional expectation of X given the current productivity p. The equilibrium is characterized by the following Bellman equations:

$$\begin{aligned} J_p &= p - w_p + \delta(1-s)\,\mathbb{E}_p\left[J_{p'}\right] && (4.2)\\ V_p &= -c_p + \delta q(\theta_p)\,\mathbb{E}_p\left[J_{p'}\right] && (4.3)\\ U_p &= z + \delta\left\{f(\theta_p)\,\mathbb{E}_p\left[W_{p'}\right] + (1 - f(\theta_p)\,\mathbb{E}_p\left[U_{p'}\right])\right\} && (4.4)\\ W_p &= w_p + \delta\left\{(1-s)\,\mathbb{E}_p\left[W_{p'}\right] + s\mathbb{E}_p\left[U_{p'}\right]\right\}. && (4.5)\end{aligned}$$

In addition, free entry implies

$$V_p = 0, \tag{4.6}$$

and Nash bargaining implies

$$J_p = (W_p - U_p)(1-\beta)/\beta. \tag{4.7}$$

From equations (4.2)–(4.7), we can solve the functions J_p, V_p, U_p, W_p and w_p in terms of θ_p, and then reduce this system of equations to the following functional equation for θ_p (see Supplemental Appendix S.B.4 for details):

$$\delta q(\theta_p)\,\mathbb{E}_p\left[(1-\beta)\left(p' - z\right) - \beta\theta_{p'}c_{p'} + (1-s)\frac{c_{p'}}{q\left(\theta_{p'}\right)}\right] - c_p = 0. \tag{4.8}$$

In standard calibration analysis, one can solve θ_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 (4.8) alternatively as a conditional moment restriction on observed data. Replacing p and θ with their observed time series yields

$$\delta q(\theta_t)\,\mathbb{E}_t\left[(1-\beta)(p_{t+1} - z) - \beta\theta_{t+1}c_{t+1} + (1-s)\frac{c_{t+1}}{q(\theta_{t+1})}\right] - c_t = 0, \tag{4.9}$$

where we write c_t in place of c_{p_t} (recall (4.1)) and use $\mathbb{E}_t$ to denote the conditional expectation given the time-t information.[21] For our discussion below, it is convenient to rewrite (4.9) equivalently as

$$\mathbb{E}_t\left[\zeta_{t+1} - z\right] = 0, \tag{4.10}$$

where the variable ζ_{t+1} does not depend on z and is defined as

$$\zeta_{t+1} \equiv p_{t+1} - \frac{\beta\theta_{t+1}c_{t+1}}{1-\beta} + \frac{(1-s)\,c_{t+1}}{(1-\beta)\,q(\theta_{t+1})} - \frac{c_t}{(1-\beta)\,\delta q(\theta_t)}. \tag{4.11}$$

Below, we refer to (4.10) as the *equilibrium conditional moment restriction* and conduct formal econometric inference based on it.

[21] Since the state process is Markovian, the time-t information set is spanned by p_t, that is, $\mathbb{E}_t\left[\,\cdot\,\right] = \mathbb{E}\left[\,\cdot\,|p_t\right]$.

4.2 Testing results for the benchmark calibration

We start with testing whether the equilibrium conditional moment restriction (4.9) holds or not at the benchmark parameter values calibrated by HM, which are summarized in Table 1.[22] It is instructive to briefly recall HM's calibration strategy. The calibration involves two stages. In the first stage, the parameters δ, s, ρ, σ_ε, c^K, c^W and ξ are calibrated by matching certain empirical quantities. These parameters are then fixed. The second stage pins down the three remaining key parameters of the model: the value of nonmarket activity z, the workers' bargaining power parameter β and the matching parameter l. These parameters are jointly determined by matching model-implied wage-productivity elasticity, average job finding rate and average market tightness with their empirical estimates.[23]

The second stage involving the nonmarket return and the bargaining parameter is the "more contentious" part of the calibration (see Hornstein, Krusell, and Violante (2005), p. 37). For this reason, we structure our investigation using the same two-stage architecture as HM; that is, we fix the first-stage parameters at their calibrated values, and intentionally focus on how the key parameters (z, β, l) interact with the equilibrium conditional moment restriction. Doing so allows us to speak directly to the core of the debate on the unemployment volatility puzzle. We are interested especially in the value of nonmarket activity z because it is the sole determinant of the fundamental surplus fraction in the standard Mortensen–Pissarides model with Nash bargaining (Ljungqvist and Sargent (2017)). For the sake of comparison, we use exactly the same data from 1951 to 2004 as in HM's analysis, where p_t and θ_t are measured using their cyclical component obtained from the Hodrick–Prescott filter with smoothing parameter 1600.[24]

The calibration described in Table 1 was conducted at the weekly frequency. Since our econometric inference is based on quarterly data, we need to adjust $(\delta, s, q(\cdot))$ accordingly to the quar-

[22] Table 1 reproduces Table 2 in Hagedorn and Manovskii (2008) except that we write $\sigma_\varepsilon = 0.0034$ instead of $\sigma_\varepsilon^2 = 0.0034$. The latter appears to be a typo, in view of the discussion in the first paragraph on p. 1695 in Hagedorn and Manovskii (2008) and their Fortran code that is available online. We also include the information for the calibrated vacancy cost function $c_p = c^K p + c^W p^\xi$ as described in Section III.B of Hagedorn and Manovskii (2008).

[23] In Section II.A of Hagedorn and Manovskii (2008), the authors state that the matching parameter l is chosen to fit the weekly average job finding rate 0.139. In Section III.C, the authors further mention that the nonmarket return z and the bargaining parameter β are chosen to fit the average labor market tightness 0.634 and the wage-productivity elasticity 0.449. In their numerical implementation, these three parameters are calibrated jointly by minimizing the sum of squared relative biases (normalized by the target values) in the wage-productivity elasticity, the average job finding rate and the average labor market tightness; see the subroutine "func" in their Fortran code that is available at the publisher's website.

[24] The data is obtained from the publisher's website. For brevity, we refer the reader to Hagedorn and Manovskii (2008) for additional information about the data.

Table 1: Calibrated Parameter Values at Weekly Frequency

Parameter	Definition	Value
z	Value of nonmarket activity	0.955
β	Workers' bargaining power	0.052
l	Matching parameter	0.407
c^K	Capital costs of posting vacancies	0.474
c^W	Labor costs of posting vacancies	0.110
ξ	Wage elasticity	0.449
δ	Discount rate	$0.99^{1/12}$
s	Separation rate	0.0081
ρ	Persistence of productivity shocks	0.9895
σ_ϵ	Standard deviation of innovations in productivity	0.0034

terly frequency (i.e., 12 weeks). We set $\delta = 0.99$. At the monthly frequency, HM estimate the job finding rate and the separation rate to be $f = 0.45$ and $\tilde{s} = 0.026$, respectively, which imply a quarterly separation rate $s = 0.047$ in the current discrete-time model.[25] The vacancy filling rate function $q(\cdot)$ at the quarterly frequency can be adjusted from its weekly counterpart as

$$q(\theta) = 1 - \left(1 - \frac{1}{\left(1+\theta^l\right)^{1/l}}\right)^{12}.$$

[25] During each month, an employed worker may be separated (not separated) from the current job, denoted S (NS), and a job seeker may find (not find) a job, denoted F (NF). We compute the quarterly separation rate s as the probability of being unemployed at the end of a quarter conditional on being employed at the beginning of the quarter. This event contains four paths: (S,F,S), (S,NF,NF), (NS,S,NF) and (NS,NS,S). Summing the probabilities along these paths, we obtain $s = \tilde{s}\left[f\tilde{s} + (1-f)^2\right] + (1-\tilde{s})\left[\tilde{s}(1-f) + (1-\tilde{s})\tilde{s}\right] \approx 0.047$. This quarterly adjustment results in a lower value than that from a "simple" adjustment $0.026 \times 3 \approx 0.078$. The latter is higher because it ignores the nontrivial possibility that the worker's employment status can switch multiple times within a quarter at the 0.45 monthly job finding rate. If this higher separation rate were used in our analysis, ζ_{t+1} would be smaller and the equilibrium moment restriction would "demand" unambiguously a lower level of nonmarket return, which in turn would generate a lower level of unemployment volatility. Hence, using the "simple" adjustment would not change qualitatively (indeed would strengthen quantitatively) our main finding on the unemployment volatility puzzle.

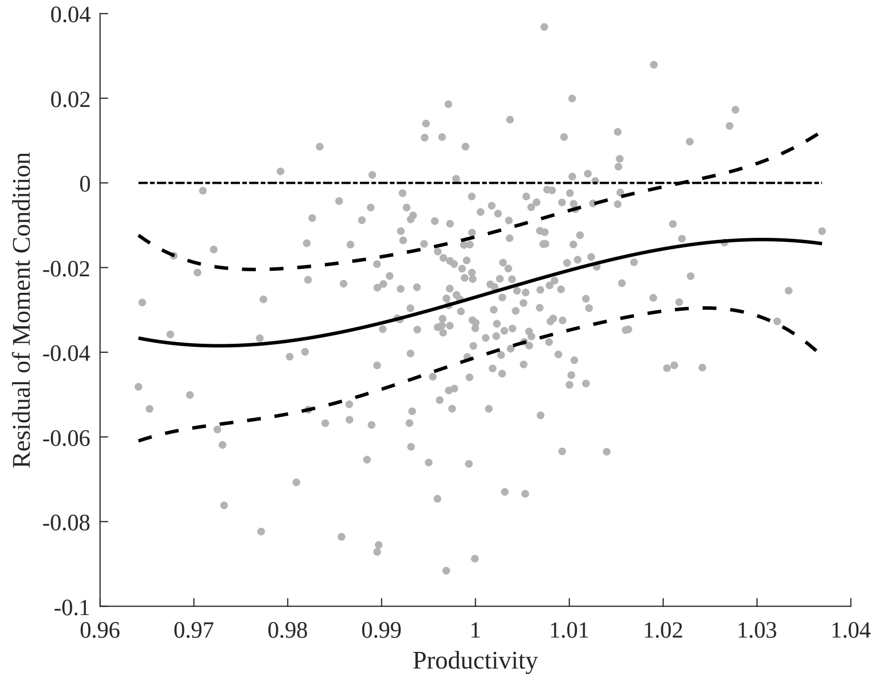

Figure 1: Nonparametric test for the conditional moment restriction under the calibration of Hagedorn and Manovskii (2008). We plot the scatter of the residual of the equilibrium conditional moment restriction $\zeta_{t+1} - z$ versus the productivity p_t, the nonparametric fit (solid) and the 95% uniform 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.

With these adjustments, we test whether the equilibrium conditional moment restriction (4.10) holds or not using the nonparametric test described in Algorithm 2 (see Section 3.5).

Figure 1 shows the scatter of the residual $\zeta_{t+1} - z$ in the moment condition (4.10) 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. By contrast, we see from the figure that the scatter of $\zeta_{t+1} - z$ is centered below zero, suggesting that the value of nonmarket activity z is too high given the other calibrated parameters. We also see a mildly positive relationship between the residual and the productivity.[26]

[26]Compared with the relatively wide uniform confidence band, the positive relation between the residual $\zeta_{t+1} - z$ and p_t is not very salient in Figure 1. That being said, this pattern alludes to the economically important notion that the value of nonmarket activity may be procyclical. Indeed, the upward sloping pattern would be conveniently "absorbed" by allowing z to be an increasing affine function of the productivity. Using detailed microdata and administrative or national accounts data, Chodorow-Reich and Karabarbounis (2016) provide direct evidence that z is procyclical and its elasticity with respect to productivity is close to one, suggesting an approximately affine rela-

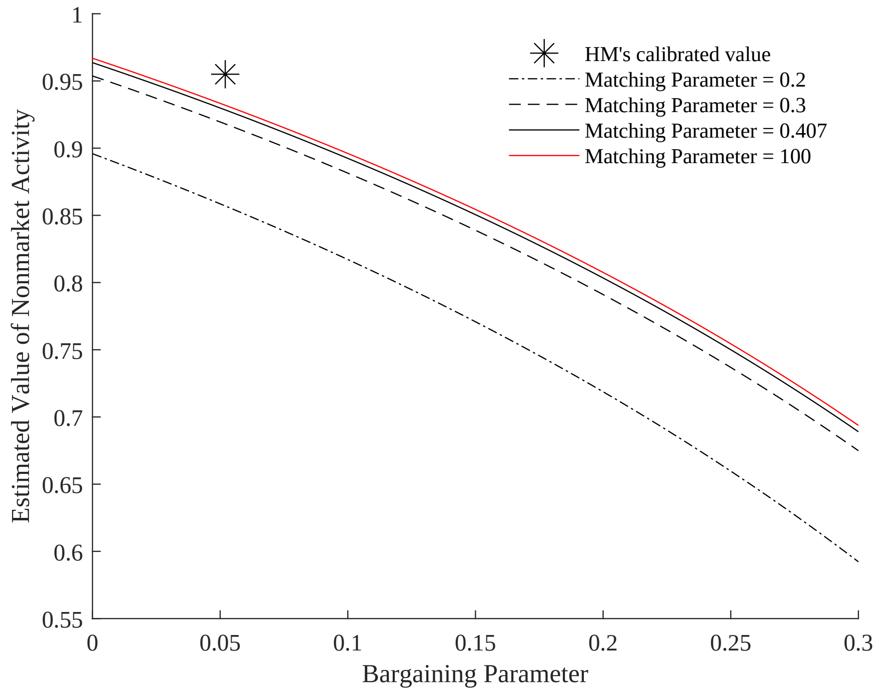

Figure 2: Empirical relationship between the value of nonmarket activity (z) and workers' bargaining power (β) implied by the unconditional moment restriction.

These patterns are more clearly revealed by the nonparametric fit of $\mathbb{E}[\zeta_{t+1} - z|p_t]$, displayed as the solid line. The uniform confidence band of the conditional moment function does not overlap with 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 parameter values in the benchmark calibration.

As clearly shown by Figure 1, the conditional moment function $\mathbb{E}[\zeta_{t+1} - z|p_t]$ would be closer to zero if a lower value of z were used; recall that the variable ζ_{t+1} does not depend on z. Varying the other parameters may serve the same purpose. In order to gain some further insight, we look at a specific *implication* of (4.10): taking an unconditional expectation on (4.10) yields

$$z = \mathbb{E}\left[\zeta_{t+1}\right]. \tag{4.12}$$

Therefore, given the other parameters, this condition may be used to uniquely determine the value of nonmarket activity z, for which the expectation in (4.12) can be estimated by a simple sample average. Figure 2 plots the point estimate of z as a function of the bargaining parameter β for various levels of the matching parameter l. We see that the estimated value of nonmarket activity

tionship between these variables. These authors also show that the procyclicality of z implies far less unemployment volatility in leading models of the labor market than that observed in the data.

is higher when the bargaining parameter β is lower and/or the matching parameter l is higher. Interestingly, increasing the matching parameter beyond 0.407 has very small effect, as illustrated by the "limiting" case with $l = 100$. Consistent with the testing result depicted in Figure 1, Figure 2 shows that the calibrated value of $z = 0.955$ (asterisk) is too high relative to the moment condition (4.12). This is true even in the limiting case with a very large value of l. The only way to obtain a 0.955 estimate for z is to change workers' bargaining power to an even lower value.

4.3 Confidence sets of admissible parameter values

We have found that the equilibrium conditional moment restriction (4.10) is rejected at the parameter values in the benchmark calibration. We now ask a more constructive question: which parameter values, if there are any, are compatible with the equilibrium conditional moment restriction (4.10)? The formal econometric answer to this question is given by the Anderson–Rubin confidence set for (z, β, l) obtained by inverting the nonparametric specification test. We remind the reader that, parallel to HM's calibration strategy, we keep the "first-stage" parameters fixed at their calibrated values and focus on the three key parameters (z, β, l).

The $(1 - \alpha)$-level Anderson–Rubin confidence set is constructed as follows. For each value of (z, β, l), we implement the nonparametric specification test for the conditional moment restriction (4.10) and include it in the confidence set if the test does not reject at the α significance level; the tests are implemented in the same way as depicted by Figure 1. To simplify the discussion, we refer to each element in the confidence set as being *admissible*. Intuitively, this confidence set embodies the econometric constraint on the parameter space implied by the equilibrium moment condition (4.10) through the lens of our test, in the sense that an admissible parameter vector satisfies this constraint statistically. As is standard in econometrics, we implement the test inversion by using a grid search: we consider $z \in [0.01, 0.99]$, $\beta \in [0.01, 0.2]$ and $l \in [0.3, 0.5]$ and discretize these intervals with mesh size 0.001, resulting in roughly 38 million grid points in total.[27]

Figure 3 shows the three-dimensional 95%-level confidence set for (z, β, l). The domain of each plot is the same as that used in the grid search. An obvious, but important, observation is that the confidence set is far away from empty.[28] In other words, the equilibrium conditional moment

[27]In this exercise, we maintain the same numerical precision for the parameters as reported in the text of Hagedorn and Manovskii (2008). Due to the computational cost, we do not consider "large" bargaining parameters or matching parameters for the following reasons. From Figure 2, we see that increasing the matching parameter has minimal effect on the equilibrium moment condition. In addition, taking large values of β would lead to lower admissible values of the nonmarket return z, which in turn will result in lower levels of unemployment volatility. Therefore, whether these (large) parameter values are included or not does not affect our main empirical findings concerning the unemployment volatility puzzle.

[28]An empty confidence set would imply a rejection of the economic model for all parameter values under consid-

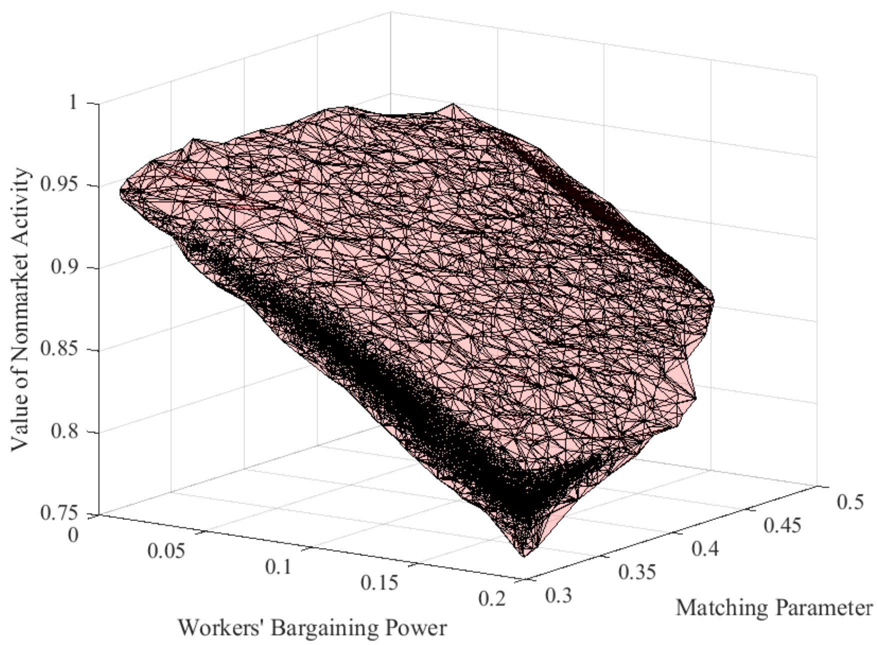

Figure 3: Anderson–Rubin confidence sets (95% level) for (z, β, l).

restriction is compatible with the data for a wide range of parameter values and, to this extent, is not overly restrictive.

We now compare HM's calibrated parameter values with those in the confidence set. To simplify the visual inspection, we plot in Figure 4 a two-dimensional slice of the 95% confidence set (light colored area) on the (z, β) plane sectioned at $l = 0.407$, which is the value of the matching parameter calibrated by HM; we also plot the 90%-level confidence set (dark colored area) for comparison. HM's calibrated values $z = 0.955$ and $\beta = 0.052$ (asterisk) appear to be "close" to the confidence sets. Indeed, adjusting (z, β) downwards to, say $(0.935, 0.052)$, would result in a "borderline" admissible parameter value on the boundary of the 95% confidence set. However, we stress that this difference is actually economically large. As HM show analytically, the tightness-productivity elasticity is proportional to $p/(p-z)$, that is, the inverse of the *fundamental surplus fraction* studied more generally by Ljungqvist and Sargent (2017). Varying z from 0.955 to 0.935 would increase the fundamental surplus fraction from 0.045 to 0.065 and subsequently reduce the tightness-productivity elasticity by roughly 30%. Therefore, this seemingly small difference in z would suppress nontrivially the model's capacity for generating volatilities in market tightness and

eration. This logic is also commonly used for specification tests in settings with weak or partial identification; see, for example, Stock and Wright (2000), Andrews and Soares (2010), among others.

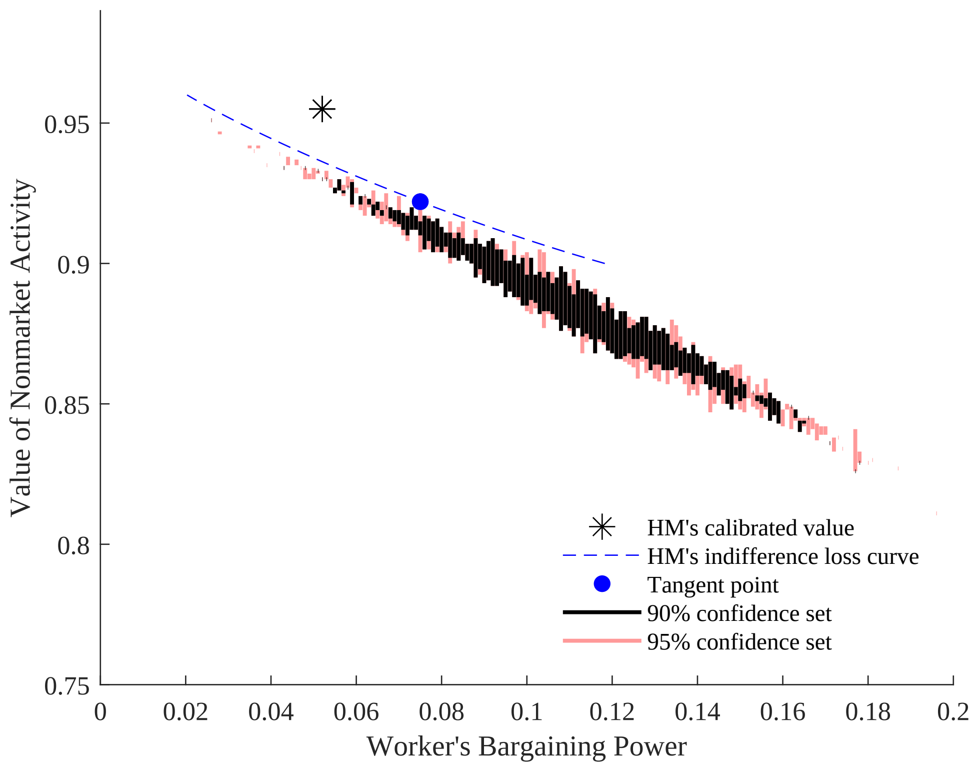

Figure 4: Two-dimensional illustration of Anderson–Rubin confidence set and constrained calibration. We plot 90% and 95% confidence sets for (z, β) obtained as slices of the corresponding three-dimensional Anderson–Rubin confidence sets sectioned at $l = 0.407$. Hagedorn and Manovskii's (2008) calibrated value is plotted for comparison, which is obtained by minimizing the loss function defined as the root mean square relative error for matching average tightness, wage-productivity elasticity and average job finding rate. The dashed line is an indifference curve induced by this loss function that is tangent to the 95% confidence set; the indifference curve is constructed from the numerical solutions of the equilibrium Bellman equations. The tangent point $(z, \beta) = (0.922, 0.075)$ depicts the solution to the constrained calibration (with l fixed at 0.407 in this illustration).

unemployment. In view of how "thin" the confidence sets in Figure 4 are in the z-dimension, this difference is also statistically highly significant, which is expected given the testing result depicted by Figure 1.

Our intention is not to suggest that the parameter values of (z, β) in the confidence sets are "better" than those calibrated by HM, because it would be difficult to make such a claim due to the lack of a "consensus" set of calibration targets. After all, HM's calibration is designed to match three important economic targets exactly, so any other choices of parameters ought to do worse in these dimensions, and vice versa. However, we do assert, based on formal econometric evidence, that the equilibrium moment condition (4.10) is violated at these parameter values at conventional significance levels. Since this condition is derived directly from the equilibrium Bellman equations, such a violation suggests a lack of internal consistency in HM's calibrated model.

How to ensure this type of internal consistency in the calibration exercise? A simple way to achieve this is to restrict the calibration among admissible parameter values in the confidence set, because the equilibrium moment condition is not rejected for those parameters by our construction. We implement this idea in two settings. In Setting (1), we choose (z, β, l) from the three-dimensional 95% confidence set depicted in Figure 4 by minimizing the same loss function as in HM, defined as the sum of squared relative calibration error relative to the wage-productivity elasticity, the average job finding rate and the average labor market tightness. In Setting (2), we impose a more stringent constraint by restricting the loss minimization within the (smaller) 90% confidence set.[29]

Using these "admissibility constrained" calibrations, we aim to seek a constructive compromise between macro-type calibration and moment-based estimation (Kydland and Prescott (1996), Hansen and Heckman (1996), Dawkins, Srinivasan, and Whalley (2001)). We bring in econometric tools: econometrically justified confidence sets are used to "discipline" the calibration. But we are not conducting GMM-type estimation: we use the calibrator's loss function, instead of an econometrician's loss function defined by instrumented moment conditions (cf. Hansen (1982)).[30]

[29] We remind the reader that, without the admissibility constraints, these calibrations would be identical to that of HM. In particular, since these parameter values are chosen to minimize the same objective function (though under different constraints), the differences in the calibrated parameters reflect exclusively the effect of the admissibility constraints. Being aware of the critiques from Hall and Milgrom (2008), Costain and Reiter (2008) and Pissarides (2009), we intentionally use the same calibration targets as HM in order to demonstrate precisely how much effect the econometric constraint has on the calibration.

[30] We note that the Anderson–Rubin type confidence set used here is very different from the confidence sets obtained from the standard GMM theory. The former is computationally more difficult to obtain (generally due to the grid search), but is immune to issues arising from weak or partial identification; see, for example, Stock and Wright (2000) for a discussion on weak identification in GMM problems.

Intuitively, GMM-type estimation would result in estimates at the "center" of the confidence sets depicted by Figures 3 and 4, which can be much different from the calibrated value because they minimize different loss functions. By contrast, the constrained calibration yields "admissible" parameter values that are the closest, formally measured by the calibration loss, to the standard (unconstrained) calibrated value. To illustrate this point graphically, we plot in Figure 4 an indifference curve (dashed) induced by HM's loss function that is tangent to the 95% confidence set. Again, we fix the matching parameter at 0.407 only for the ease of visual illustration. The tangent point between the indifference curve and the confidence set corresponds to the parameter value obtained from a constrained calibration. The standard GMM estimate, on the other hand, is at the center of the confidence set and is evidently further away from HM's calibrated values than the tangent point.

We now turn to the results. Table 2 compares the calibrated parameter values of (z, β, l) in these two constrained calibrations with HM's unconstrained benchmark. Not surprisingly, since the parameters are chosen to minimize the same loss function, they appear to be "numerically close" to each other. However, as emphasized above, the seemingly small differences in the nonmarket return z actually correspond to large differences in the fundamental surplus fraction $(p-z)/p$ and, hence, are economically significant. In Panel A of Table 3, we report the model-implied values of the target variables and compare them with the empirical estimates. The fourth row summarizes the goodness of fit, defined as the root mean square relative calibration error for matching the three calibration targets. As expected, HM's calibrated parameters result in an almost exact fitting.[31] Under the admissibility constraint, the relative calibration error is 13% in Setting (1) and is 17% in the more constrained Setting (2); intuitively, these numbers gauge the "cost" a calibrator needs to pay for satisfying statistically the equilibrium conditional moment restriction.

We see from Table 2 that the value of nonmarket activity in the constrained calibrations are lower than that in the unconstrained benchmark. Hence, in theory, we expect the unemployment volatility generated in the former to be lower than that in the latter. To anticipate how much the effect is, we can do some back-of-the-envelope calculations using the theory of Ljungqvist and Sargent (2017). Note that reducing z from 0.955 to 0.941 (resp. 0.926) decreases the inverse of fundamental surplus fraction by roughly 24% (resp. 39%). As a coarse theoretical approximation, we expect the unemployment volatility to drop by a similar amount.

To quantify this effect precisely, we solve the equilibrium numerically in each calibration set-

[31] The calibration error using HM's calibrated values is slightly larger than that reported in their paper. The reason for this small difference is that we use the parameter values reported in their main text, which are less precise than those actually used in their numerical work.

Table 2: Calibrated Parameter Values at Weekly Frequency

		Calibration Setting		
Parameter	Definition	HM	(1)	(2)
z	Value of nonmarket activity	0.955	0.941	0.926
β	Workers' bargaining power	0.052	0.058	0.077
l	Matching parameter	0.407	0.373	0.337

Table 3: Simulation Results for Calibrated Models

		Calibration Setting		
Variable	Data	HM	(1)	(2)
Panel A. Calibration Targets				
Wage-productivity elasticity	0.449	0.451	0.459	0.486
Average job finding rate	0.139	0.140	0.134	0.108
Average market tightness	0.634	0.644	0.772	0.737
Relative calibration error	–	1%	13%	17%
Panel B. Volatility Measures				
Unemployment	0.125	0.143	0.093	0.068
Vacancy	0.139	0.167	0.117	0.093
Tightness	0.259	0.289	0.198	0.150

ting, and then compute the volatility of unemployment, vacancy and labor market tightness by simulation.[32] Panel B of Table 3 reports the simulated volatilities along with the corresponding empirical estimates. From the second column, we observe that HM's calibrated values indeed generate sizable volatilities in unemployment and labor market tightness, which are actually higher than the empirical estimates reported in the first column; this is consistent with HM's finding that their alternative calibration strategy can be used to address the unemployment volatility puzzle in the standard Mortensen–Pissarides model. However, when we re-calibrate (z, β, l) under the admissibility constraint in Setting (1), the simulated volatilities are reduced roughly by 35% and is 26% lower than the empirical estimate. This difference becomes even larger if the restrict the calibration within the 90%-level confidence set (Setting (2)). In this case, the simulated unemployment volatility is 52% lower than that in HM and is 46% lower than the empirical estimate. These numerical findings are in line with our back-of-the-envelope theoretical calculation. They suggest that a notable portion of the observed unemployment volatility is not explained by the Mortensen–Pissarides model once we "discipline" the calibration by imposing the admissibility constraint depicted by the confidence sets. The Shimer critique is clearly in force, although to a smaller extent in our exercise (cf. Shimer (2005)).

4.4 Discussions on the unemployment volatility puzzle

We summarize our empirical findings as follows. First, our nonparametric specification test provides strong evidence that HM's calibrated parameter values are incompatible with the equilibrium conditional moment restriction directly implied by the equilibrium Bellman equations. Constructively, we compute the Anderson–Rubin confidence set consisting of admissible values of the three key parameters (z, β, l) which, by construction, satisfy statistically the equilibrium condition through the lens of our specification test. The confidence set contains a wide range of parameter values, generally with lower nonmarket return than the benchmark value $z = 0.955$. We revisit HM's calibration strategy used for addressing the unemployment volatility puzzle. We follow the same calibration strategy, except that we require the parameters to be compatible with the equilibrium conditional moment restriction (4.10) by imposing the admissibility constraint. As a result, the model-implied unemployment volatility is 26%–46% lower than the empirical estimate, depending on the choice of confidence level. Although the volatilities in these calibrations

[32]We use the same algorithm as Hagedorn and Manovskii (2008) that is available from the publisher's website. The model is solved at the weekly frequency. The simulated data are then aggregated to the quarterly frequency. We extract the cyclical components via the Hodrick–Prescott filter with smoothing parameter 1600, and then report their standard deviations as the volatility measures.

are much higher than those generated by the "common" calibration (cf. Shimer (2005)), they are still notably lower than the observed volatility of unemployment.

These empirical findings shed some light on the debate concerning the unemployment volatility puzzle. Several authors have argued that the $z = 0.955$ value of nonmarket activity calibrated by HM may be too high according to various metrics, such as the labor supply elasticity or the effect of unemployment policy; see, for example, Hall and Milgrom (2008), Costain and Reiter (2008) and Pissarides (2009). These authors, among others, thus assert the necessity for modifying the standard Mortensen–Pissarides model in various ways (e.g., by introducing wage stickiness or fixed matching cost) so as to confront the Shimer critique. Our argument is complementary to, but conceptually distinct from, these papers: we examine the "internal consistency" of the calibrated model by formally testing the equilibrium conditional moment restriction directly derived from the Bellman equations, while being agnostic about which "external" calibration targets should be picked for evaluating a competing set of calibrated parameters. Once we restrict the calibration within admissible parameter values, we find that the unemployment volatility puzzle is still present to a nontrivial extent, despite that we choose exactly the same calibration targets as HM. Our findings thus suggest that the unemployment volatility puzzle cannot be completely addressed by using HM's alternative calibration strategy, even if one would be willing to ignore other important criteria such as those emphasized by Hall and Milgrom (2008), Costain and Reiter (2008) and Pissarides (2009). Hence, we conclude that modifying the standard Mortensen–Pissarides model using insight from Shimer (2004), Hall (2005), Mortensen and Nagypál (2007), Hornstein, Krusell, and Violante (2007), Hall and Milgrom (2008), Gertler and Trigari (2009), Pissarides (2009), among others, is necessary for a better understanding of the cyclical behavior of unemployment; see Ljungqvist and Sargent (2017) for a more complete list of contributions. In principle, our econometric methodology can also be applied to examine these alternative models, but this task is beyond the scope of the present paper.

5 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. Inspired by the recent work of Chernozhukov, Lee, and Rosen (2013) and Belloni, Chernozhukov, Chetverikov, and Kato (2015), we develop a uniform inference theory for series estimators, relying crucially on our novel strong approximation theory for heterogeneous dependent data with growing dimensions. To conduct feasible inference, we also extend the classical HAC estimation theory to a high-dimensional setting. 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.

To demonstrate empirically the usefulness of our theory, we apply the proposed inference procedure to the classical Mortensen–Pissarides search and matching model for equilibrium unemployment, with a special focus on the unemployment volatility puzzle. The question about what are the "plausible" parameter values is at the center of this debate. We contribute to this literature by resorting to standard econometric principles: we derive an equilibrium conditional moment restriction from the Bellman equations and nonparametrically test its validity for a broad range of parameters. We reject Hagedorn and Manovskii's (2008) calibrated values, and find that a constrained version of their calibration subject to an "admissibility" requirement (implied by our test) would lead to an economically nontrivial difference in the key model parameters, resulting in 26%–46% of unexplained unemployment volatility. Our findings suggest that Hagedorn and Manovskii's alternative calibration strategy cannot address completely the unemployment volatility puzzle and, hence, confirm the necessity of modifying the standard search and matching model as is done in the recent macroeconomics literature.

References

Aït-Sahalia, Y., and J. Jacod (2014): *High-Frequency Financial Econometrics*. Princeton University Press.

Andersen, T. G. (1996): "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *Journal of Finance*, 51, 169–204.

Andrews, D. W. K. (1991a): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59(2), 307–345.

——— (1991b): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015): "Some new asymptotic theory for least squares series: Pointwise and uniform results," *Journal of Econometrics*, 186(2), 345 – 366, High Dimensional Problems in Econometrics.

Bierens, H. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.

Bierens, H. J., and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65(5), pp. 1129–1151.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31(3), 307 – 327.

Bollerslev, T., J. Li, and Y. Xue (2018): "Volume, Volatility and Public News Announcements," *Review of Economic Studies, forthcoming.*

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, chap. 76. Elsevier, 1 edn.

Chen, X., and T. M. Christensen (2015): "Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions," *Journal of Econometrics*, 188(2), 447–465.

——— (2018): "Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression," *Quantitative Economics*, 9(1), 39–84.

Chen, X., Z. Liao, and Y. Sun (2014): "Sieve inference on possibly misspecified semi-nonparametric time series models," *Journal of Econometrics*, 178, 639–658.

Chen, X., and D. Pouzo (2012): "Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals," *Econometrica*, 80(1), 277–321.

——— (2015): "Sieve Wald and QLR inferences on semi/nonparametric conditional moment models," *Econometrica*, 83(3), 1013–1079.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013): "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors," *Ann. Statist.*, 41(6), 2786–2819.

——— (2014): "Gaussian approximation of suprema of empirical processes," *Ann. Statist.*, 42(4), 1564–1597.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

Chodorow-Reich, G., and L. Karabarbounis (2016): "The Cyclicality of the Opportunity Cost of Employment," *Journal of Political Economy*, 124(6), 1563–1618.

Clark, P. K. (1973): "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica*, 41(1), 135–155.

Costain, J. S., and M. Reiter (2008): "Business Cycles, Unemployment Insurance, and the Calibration of Matching Models," *Journal of Economic Dynamics and Control*, 32, 1120–1155.

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford University Press.

Dawkins, C., T. Srinivasan, and J. Whalley (2001): "Chapter 58 - Calibration," vol. 5 of *Handbook of Econometrics*, pp. 3653 – 3703. Elsevier.

De Jong, R. M. (2000): "A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance matrix estimators," *Econometric Theory*, 16(2), 262–268.

de Jong, R. M., and J. Davidson (2000): "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," *Econometrica*, 68(2), 407–423.

den Haan, W. J., G. Ramey, and J. Watson (2000): "Job Destruction and Propagation of Shocks," *The American Economic Review*, 90(3), 482–498.

Eastwood, B. J. (1991): "Asymptotic normality and consistency of semi-nonparametric regression estimators using an upwards F test truncation rule," *Journal of Econometrics*, 48(1), 151 – 181.

Engle, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987–1008.

Engle, R. F., and V. K. Ng (1993): "Measuring and Testing the Impact of News on Volatility," *The Journal of Finance*, 48, 1749–1778.

Gallant, A., and G. Souza (1991): "On the asymptotic normality of Fourier flexible form estimates," *Journal of Econometrics*, 50(3), 329 – 353.

Gallant, A. R., P. E. Rossi, and G. Tauchen (1992): "Stock Prices and Volume," *The Review of Financial Studies*, 5(2), 199–242.

——— (1993): "Nonlinear Dynamic Structures," *Econometrica*, 61(4), 871–907.

Gertler, M., and A. Trigari (2009): "Unemployment Fluctuations with Staggered Nash Wage Bargaining," *Journal of Political Economy*, 117(1), 38–86.

Hagedorn, M., and I. Manovskii (2008): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited," *The American Economic Review*, 98(4), 1692–1706.

Hall, P., and C. Heyde (1980): "Martingale Limit Theory and its Application," Probability and Mathematical Statistics: A Series of Monographs and Textbooks. Academic Press.

Hall, R. E. (2005): "Employment Fluctuations with Equilibrium Wage Stickiness," *The American Economic Review*, 95(1), 50–65.

Hall, R. E., and P. R. Milgrom (2008): "The Limited Influence of Unemployment on the Wage Bargain," *The American Economic Review*, 98(4), 1653–1674.

Hannan, E. J. (1970): *Multiple Time Series*. Wiley, New York.

Hansen, B. (2015): "A Unified Asymptotic Distribution Theory for Parametric and Non- Parametric Least Squares," *Working paper*.

Hansen, B. E. (1992): “Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes,” *Econometrica*, 60(4), 967–972.

Hansen, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029–1054.

Hansen, L. P., and J. J. Heckman (1996): “The Empirical Foundations of Calibration,” *The Journal of Economic Perspectives*, 10(1), 87–104.

Hansen, P. R. (2005): “A test for superior predictive ability,” *Journal of Business & Economic Statistics*, 23(4), 365–380.

Hansen, P. R., A. Lunde, and J. M. Nason (2011): “The model confidence set,” *Econometrica*, 79(2), 453–497.

Hardle, W., and E. Mammen (1993): “Comparing Nonparametric Versus Parametric Regression Fits,” *Ann. Statist.*, 21(4), 1926–1947.

Hornstein, A., P. Krusell, and G. Violante (2005): “Unemployment and Vacancy Flucutations in the Maching Model: Inspecting the Mechanism,” *Rederal Reserve Bank Richmond Economic Quarterly*, 91(3), 19–51.

Hornstein, A., P. Krusell, and G. L. Violante (2007): “Technology-Policy Interaction in Frictional Labour-Markets,” *The Review of Economic Studies*, 74(4), 1089–1124.

Jacod, J., and P. Protter (2012): *Discretization of Processes*. Springer.

Jacod, J., and A. N. Shiryaev (2003): *Limit Theorems for Stochastic Processes*. Springer-Verlag, New York, second edn.

Kandel, E., and N. D. Pearson (1995): “Differential Interpretation of Public Signals and Trade in Speculative Markets,” *Journal of Political Economy*, 103(4), pp. 831–872.

Karpoff, J. M. (1987): “The Relation between Price Changes and Trading Volume: A Survey,” *Journal of Financial and Quantitative Analysis*, 22(1), 109–126.

Kim, O., and R. E. Verrecchia (1991): “Trading Volume and Price Reactions to Public Announcements,” *Journal of Accounting Research*, 29(2), 302–321.

Kydland, F. E., and E. C. Prescott (1996): “The Computational Experiment: An Econometric Tool,” *The Journal of Economic Perspectives*, 10(1), 69–85.

Kyle, A. S. (1985): “Continuous Auctions and Insider Trading,” *Econometrica*, 53(6), 1315–1335.

Ljungqvist, L., and T. J. Sargent (2012): *Recursive Macroeconomic Theory*. MIT Press.

——— (2017): “The Fundamental Surplus,” *American Economic Review*, 107(9), 2630–2665.

Mikusheva, A. (2007): “Uniform Inference in Autoregressive Models,” *Econometrica*, 75(5), 1411–1452.

Mortensen, D. T., and E. Nagypál (2007): "More on Unemployment and Vacancy Fluctuations," *Review of Economic Dynamics*, 10, 327–347.

Mortensen, D. T., and C. A. Pissarides (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *The Review of Economic Studies*, 61(3), 397–415.

Nelson, D. B. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59(2), 347–370.

Newey, W. K. (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79(1), 147 – 168.

Newey, W. K., and D. L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, vol. 4. Elsevier.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

Pissarides, C. (2000): *Equilibrium Unemployment Theory*. MIT Press, Cambridge, MA.

Pissarides, C. A. (1985): "Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages," *The American Economic Review*, 75(4), 676–690.

——— (2009): "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?," *Econometrica*, 77(5), 1339–1369.

Pollard, D. (2001): *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.

Romano, J. P., and M. Wolf (2005): "Stepwise multiple testing as formalized data snooping," *Econometrica*, 73(4), 1237–1282.

Shimer, R. (2004): "The Consequences of Rigid Wages in Search Models," *Journal of the European Economic Association*, 2(2/3), 469–479.

——— (2005): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *The American Economic Review*, 95(1), 25–49.

Stock, J. H., and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68(5), 1055–1096.

Tauchen, G., H. Zhang, and M. Liu (1996): "Volume, volatility, and leverage: A dynamic analysis," *Journal of Econometrics*, 74(1), 177 – 208.

Tauchen, G. E., and M. Pitts (1983): "The Price Variability-volume Relationship on Speculative Markets," *Econometrica*, 51(2), 485–505.

van de Geer, S. (1990): "Estimating a Regression Function," *Ann. Statist.*, 18(2), 907–924.

van der Vaart, A., and J. Wellner (1996): *Weak Convergence and Empirical Processes.* Springer-Verlag.

White, H. (2000): "A reality check for data snooping," *Econometrica*, 68(5), 1097–1126.

White, H. (2001): *Asymptotic Theory for Econometricians.* Academic Press.

Wu, W. B. (2005): "Nonlinear system theory: Another look at dependence," *Proceedings of the National Academy of Sciences*, 102(40), 14150–14154.

Yurinskii, V. V. (1978): "On the Error of the Gaussian Approximation for Convolutions," *Theory of Probability & Its Applications*, 22(2), 236–247.

Zhang, D., and W. B. Wu (2017): "Gaussian approximation for high dimensional time series," *Ann. Statist.*, 45(5), 1895–1919.

Supplemental Appendix to
Uniform Nonparametric Series Inference for Dependent Data with an Application to the Search and Matching Model

Jia Li* and Zhipeng Liao†

June 17, 2018

Abstract

This technical appendix contains two sections. Section S.A contains the proofs for all results in the main text. Section S.B provides some additional results that are mentioned in the main text.

*Department of Economics, Duke University, Durham, NC 27708; e-mail: jl410@duke.edu.
†Department of Economics, UCLA, Log Angeles, CA 90095; e-mail: zhipeng.liao@econ.ecla.edu.

S.A 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 $\mathrm{Tr}(A)$, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively. We use $a^{(j)}$ to denote the jth 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 Proposition 1

Proof of Proposition 1. We observe that

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

We then prove the assertions for the α-mixing and ϕ-mixing cases separately.

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

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

Therefore, we can further bound the terms in (A.1) as follows

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

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

The ϕ-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),

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

From here, the assertion of the proposition for the ϕ-mixing case readily follows. *Q.E.D.*

S.A.2 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 Σ_n; see Lemma A1. We then establish the coupling between S_n^* and $\widetilde{S}_n$ 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 τ_n is a stopping time because $V_{n,t}$ is $\mathcal{F}_{n,t-1}$-measurable for each t and Σ_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 τ_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

$$V_{n,t}^* = \mathbb{E}\left[Z_{n,t} Z_{n,t}^\top \middle| \mathcal{H}_{n,t-1} \right] \tag{A.2}$$

and set

$$S_n^* = \sum_{t=1}^{k_n+K_n} Z_{n,t}. \tag{A.3}$$

The conditional covariance matrix of S_n^* is exactly Σ_n, that is,

$$\sum_{t=1}^{k_n+K_n} V_{n,t}^* = \sum_{t=1}^{\tau_n} V_{n,t} + \xi_n = \Sigma_n. \tag{A.4}$$

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 $u_n^* = \lceil C_1 r_n k_n \rceil$ and $h_n^* = k_n - u_n^*$,

$$\limsup_{n\to\infty} \mathbb{P}\left(\tau_n < h_n^*\right) < \varepsilon. \tag{A.5}$$

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 $h_n/k_n \to 1$,

$$\limsup_{n\to\infty} \mathbb{P}\left(\lambda_{\max}\left(\sum_{t=1}^{h_n} V_{n,t} - \Sigma_{n,h_n}\right) > C_2 r_n\right) < \varepsilon. \tag{A.6}$$

Let $\underline{\lambda} > 0$ denote a lower bound for the eigenvalues as described in Assumption 1(i). We shall show that (A.5) holds for $C_1 \equiv C_2/\underline{\lambda}$.

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.6) holds for $h_n = h_n^*$. Moreover, observe that

$$\frac{u_n^*}{r_n k_n} = \frac{\lceil C_1 r_n k_n \rceil}{r_n k_n} \geq C_1 = \frac{C_2}{\underline{\lambda}},$$

which, together with the definition of $\underline{\lambda}$, implies that

$$C_2 r_n \leq \frac{u_n^*}{k_n}\underline{\lambda} \leq \lambda_{\min}\left(\sum_{t=h_n^*+1}^{k_n} \mathbb{E}\left[V_{n,t}\right]\right). \tag{A.7}$$

We then observe

$$\begin{aligned}
\mathbb{P}\left(\tau_n < h_n^*\right) &\leq \mathbb{P}\left(\lambda_{\max}\left(\sum_{t=1}^{h_n^*} V_{n,t} - \Sigma_n\right) > 0\right) \\
&= \mathbb{P}\left(\lambda_{\max}\left(\sum_{t=1}^{h_n^*} V_{n,t} - \Sigma_{n,h_n^*} - (\Sigma_n - \Sigma_{n,h_n^*})\right) > 0\right) \\
&\leq \mathbb{P}\left(\lambda_{\max}\left(\sum_{t=1}^{h_n^*} V_{n,t} - \Sigma_{n,h_n^*}\right) > \lambda_{\min}\left(\sum_{t=h_n^*+1}^{k_n} \mathbb{E}\left[V_{n,t}\right]\right)\right) \\
&\leq \mathbb{P}\left(\lambda_{\max}\left(\sum_{t=1}^{h_n^*} V_{n,t} - \Sigma_{n,h_n^*}\right) > C_2 r_n\right),
\end{aligned} \tag{A.8}$$

where the first inequality follows from the definition of τ_n, the second inequality follows from the property of eigenvalues and the last inequality is by (A.7). From (A.6) and (A.8), the claim (A.5) 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

$$\sum_{t=1}^{k_n} X_{n,t} 1_{\{t>\tau_n\}} = O_p(m_n^{1/2} r_n^{1/2}), \quad K_n^{-1/2}\xi_n^{1/2} \sum_{t=k_n+1}^{k_n+K_n} \eta_{n,t} = O_p(m_n^{1/2} r_n^{1/2}). \tag{A.9}$$

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 τ_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

$$\begin{aligned} \mathbb{E}\left[\left\|\sum_{t=1}^{k_n} X_{n,t} 1_{\{t>\tau_n\}}\right\|^2\right] &= \mathbb{E}\left[\sum_{t=1}^{k_n} \|X_{n,t}\|^2 1_{\{t>\tau_n\}}\right] \\ &\leq \sum_{t=h_n^*+1}^{k_n} \mathbb{E}\left[\|X_{n,t}\|^2\right] = \mathrm{Tr}\left[\sum_{t=h_n^*+1}^{k_n} \mathbb{E}\left[V_{n,t}\right]\right]. \end{aligned}$$

By Assumption 1(i), the majorant side of the above inequality is $O\left(u_n^* m_n / k_n\right) = O\left(m_n r_n\right)$. The first assertion in (A.9) then readily follows.

Turning to the second assertion in (A.9), we note that

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

By the same argument as above, the majorant side of the above inequality is $O\left(m_n r_n\right)$, which implies the second assertion in (A.9). *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 Σ_n. Suppose that $K_n \geq 36 m_n^3 \bar{\lambda}^3 / B_n^2$. Then, there exists a sequence $\widetilde{S}_n$ of m_n-dimensional centered Gaussian random vectors with covariance matrix Σ_n such that*

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

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

$$\frac{\delta_{C,n}^2}{m_n \sigma_n^2} = C^2 \text{ and } \frac{B_n}{\sigma_n^2 \delta_{C,n}} = C^{-1}. \tag{A.10}$$

Below, we denote

$$\psi_{C,n} \equiv \left(\frac{C^2}{\exp\left(C^2 - 1\right)} \right)^{m_n/2}.$$

Note that as $C \to \infty$,

$$\psi_{C,n} \to 0 \text{ uniformly in } n. \tag{A.11}$$

In step 2, below, we show that the following inequality holds for any Borel subset $A \subseteq \mathbb{R}^{m_n}$:

$$\mathbb{P}\left(S_n^* \in A\right) \le F_n\left(A^{3\delta_{C,n}}\right) + \frac{1}{1 - \psi_{C,n}} \left(\psi_{C,n} + \frac{4B_n}{\sigma_n^2 \delta_{C,n}} \right), \tag{A.12}$$

where F_n denotes the distribution of an $\mathcal{N}(0, \Sigma_n)$ random variable and

$$A^{3\delta_{C,n}} \equiv \left\{ x \in \mathbb{R}^{m_n} : \inf_{y \in A} \|x - y\| \le 3\delta_{C,n} \right\}.$$

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

$$\begin{aligned} \mathbb{P}\left(\left\| S_n^* - \widetilde{S}_n \right\| > 3\delta_{C,n} \right) &\le \frac{1}{1 - \psi_{C,n}} \left(\psi_{C,n} + \frac{4B_n}{\sigma_n^2 \delta_{C,n}} \right) \\ &= \frac{1}{1 - \psi_{C,n}} \left(\psi_{C,n} + 4C^{-1} \right). \end{aligned}$$

By (A.11), 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 ε, yielding

$$\mathbb{P}\left(\left\| S_n^* - \widetilde{S}_n \right\| > 3C(B_n m_n)^{1/3} \right) < \varepsilon.$$

This proves the assertion of the lemma.

Step 2. It remains to show (A.12). For notational simplicity, we write δ_n and ψ_n in place of $\delta_{C,n}$ and $\psi_{C,n}$, respectively. With σ_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 $d\left(x, A^{\delta_n}\right)$ denotes the distance between x and the set A^{δ_n}. By Lemma 10.18 in Pollard (2001), $f_n(\cdot)$ is three-time continuously differentiable such that for all (x, y),

$$\left| f_n(x + y) - f_n(x) - \partial f_n(x)^\top y - \frac{1}{2} y^\top \partial^2 f_n(x) y \right| \le \frac{\|y\|^3}{\sigma_n^2 \delta_n}, \tag{A.13}$$

and

$$(1-\psi_n)1\{x\in A\}\le f_n(x)\le \psi_n+(1-\psi_n)1\left\{x\in A^{3\delta_n}\right\}. \tag{A.14}$$

Let $\zeta_{n,t}$, $1\le t\le k_n+K_n$, be independent m_n-dimensional standard normal vectors and $\tilde{\zeta}_{n,t}\equiv (V_{n,t}^*)^{1/2}\zeta_{n,t}$; recall the definition of $V_{n,t}^*$ from (A.2). We set

$$D_{n,t}\equiv \sum_{1\le s<t} Z_{n,s}+\sum_{t<s\le k_n+K_n}\tilde{\zeta}_{n,s}.$$

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

$$D_{n,t}+Z_{n,t}=D_{n,t+1}+\tilde{\zeta}_{n,t+1},\quad 1\le t\le k_n+K_n-1.$$

Hence,

$$\mathbb{E}\left[f_n(S_n^*)\right]-\int f_n(x)F_n(dx)=\sum_{t=1}^{k_n+K_n}\left(\mathbb{E}\left[f_n(D_{n,t}+Z_{n,t})\right]-\mathbb{E}\left[f_n(D_{n,t}+\tilde{\zeta}_{n,t})\right]\right). \tag{A.15}$$

By (A.13), we have

$$\begin{aligned}&\Big|\mathbb{E}\left[f_n(D_{n,t}+Z_{n,t})\right]-\mathbb{E}\left[f_n(D_{n,t})\right]-\mathbb{E}[\partial f_n(D_{n,t})^\top Z_{n,t}]\\ &\qquad -(1/2)\mathbb{E}[\mathrm{Tr}(\partial^2 f_n(D_{n,t})Z_{n,t}Z_{n,t}^\top)]\Big|\le \frac{1}{\sigma_n^2\delta_n}\mathbb{E}[\,\|Z_{n,t}\|^3],\end{aligned} \tag{A.16}$$

and

$$\begin{aligned}&\Big|\mathbb{E}[f_n(D_{n,t}+\tilde{\zeta}_{n,t}]-\mathbb{E}\left[f_n(D_{n,t})\right]-\mathbb{E}[\partial f_n(D_{n,t})^\top\tilde{\zeta}_{n,t}]\\ &\qquad -(1/2)\mathbb{E}[\mathrm{Tr}(\partial^2 f_n(D_{n,t})\tilde{\zeta}_{n,t}\tilde{\zeta}_{n,t}^\top)]\Big|\le \frac{1}{\sigma_n^2\delta_n}\mathbb{E}[\|\tilde{\zeta}_{n,t}\|^3].\end{aligned} \tag{A.17}$$

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

$$\mathbb{E}[\partial f_n(D_{n,t})^\top\tilde{\zeta}_{n,t}]=0 \text{ and } \mathbb{E}[\mathrm{Tr}(\partial^2 f_n(D_{n,t})\tilde{\zeta}_{n,t}\tilde{\zeta}_{n,t}^\top)]=\mathbb{E}[\mathrm{Tr}(\partial^2 f_n(D_{n,t})V_{n,t}^*)]. \tag{A.18}$$

Let $\widetilde{D}_{n,t}\equiv\sum_{1\le s<t}Z_{n,s}+(\Sigma_n-\sum_{s=1}^t V_{n,t}^*)\zeta_{n,t}$. We note that since Σ_n is nonrandom, $\Sigma_n-\sum_{s=1}^t V_{n,t}^*$ is $\mathcal{H}_{n,t-1}$-measurable. We then observe that

$$\begin{aligned}\mathbb{E}[\partial f_n(D_{n,t})^\top Z_{n,t}] &= \mathbb{E}[\partial f_n(\widetilde{D}_{n,t})^\top Z_{n,t}]\\ &= \mathbb{E}[\partial f_n(\widetilde{D}_{n,t})^\top\mathbb{E}[Z_{n,t}|\mathcal{H}_{n,t-1},\zeta_{n,t}]]\\ &= \mathbb{E}[\partial f_n(\widetilde{D}_{n,t})^\top\mathbb{E}[Z_{n,t}|\mathcal{H}_{n,t-1}]]=0,\end{aligned} \tag{A.19}$$

where the first equality holds because the conditional distribution of $\widetilde{D}_{n,t}$ given $\mathcal{H}_{n,k_n+K_n}$ is the same as that of $D_{n,t}$; the second equality holds because $\sum_{1\le s<t}Z_{n,s}$ and $\Sigma_n-\sum_{s=1}^t V_{n,t}^*$ are

$\mathcal{H}_{n,t-1}$-measurable; the third equality is by the independence between $\zeta_{n,t}$ and $\mathcal{H}_{n,k_n+K_n}$ and the last equality holds because $(Z_{n,t}, \mathcal{H}_{n,t})_{1\leq t\leq k_n+K_n}$ is a martingale difference array by construction. Similarly,

$$\begin{aligned}
\mathbb{E}[\mathrm{Tr}(\partial^2 f_n(D_{n,t})Z_{n,t}Z_{n,t}^\top)] &= \mathbb{E}[\mathrm{Tr}(\partial^2 f_n(\widetilde{D}_{n,t})Z_{n,t}Z_{n,t}^\top)] \\
&= \mathbb{E}[\mathrm{Tr}(\partial^2 f_n(\widetilde{D}_{n,t})\mathbb{E}[Z_{n,t}Z_{n,t}^\top|\mathcal{H}_{n,t-1},\zeta_{n,t}])] \\
&= \mathbb{E}[\mathrm{Tr}(\partial^2 f_n(\widetilde{D}_{n,t})\mathbb{E}[Z_{n,t}Z_{n,t}^\top|\mathcal{H}_{n,t-1}])] \\
&= \mathbb{E}[\mathrm{Tr}(\partial^2 f_n(\widetilde{D}_{n,t})V_{n,t}^*)] = \mathbb{E}[\mathrm{Tr}(\partial^2 f_n(D_{n,t})V_{n,t}^*)]. \qquad \text{(A.20)}
\end{aligned}$$

Combining the results in (A.18), (A.19) and (A.20), we have

$$\begin{aligned}
\mathbb{E}[\partial f_n(D_{n,t})^\top Z_{n,t}] &= \mathbb{E}[\partial f_n(D_{n,t})^\top \tilde{\zeta}_{n,t}] = 0 \\
\mathbb{E}[\mathrm{Tr}(\partial^2 f_n(D_{n,t})Z_{n,t}Z_{n,t}^\top)] &= \mathbb{E}[\mathrm{Tr}(\partial^2 f_n(D_{n,t})\tilde{\zeta}_{n,t}\tilde{\zeta}_{n,t}^\top)].
\end{aligned}$$

Combining this with (A.15), (A.16) and (A.17), we deduce

$$\begin{aligned}
&\left|\mathbb{E}\left[f(S_n^*)\right] - \int f_n(x)F_n(dx)\right| \\
&\leq \frac{1}{\sigma_n^2\delta_n}\sum_{t=1}^{k_n+K_n}\left(\mathbb{E}[\,\|Z_{n,t}\|^3] + \mathbb{E}[\|\tilde{\zeta}_{n,t}\|^3]\right) \\
&= \frac{1}{\sigma_n^2\delta_n}\sum_{t=1}^{k_n}\left(\mathbb{E}[\,\left\|X_{n,t}1_{\{t\leq\tau_n\}}\right\|^3] + \mathbb{E}[\|V_{n,t}^{1/2}\zeta_{n,t}1_{\{t\leq\tau_n\}}\|^3]\right) + \frac{2}{\sigma_n^2\delta_n}\sum_{t=k_n+1}^{k_n+K_n}\mathbb{E}\left[\left\|K_n^{-1/2}\xi_n^{1/2}\eta_{n,t}\right\|^3\right] \\
&\leq \frac{1}{\sigma_n^2\delta_n}\sum_{t=1}^{k_n}\left(\mathbb{E}[\,\|X_{n,t}\|^3] + \mathbb{E}[\|V_{n,t}^{1/2}\zeta_{n,t}\|^3]\right) + \frac{2}{\sigma_n^2\delta_n}\sum_{t=k_n+1}^{k_n+K_n}\mathbb{E}\left[\left\|K_n^{-1/2}\xi_n^{1/2}\eta_{n,t}\right\|^3\right] \\
&\leq \frac{3B_n}{\sigma_n^2\delta_n} + \frac{2}{\sigma_n^2\delta_n K_n^{1/2}}\mathbb{E}\left[\left\|\xi_n^{1/2}\mathcal{N}\right\|^3\right],
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\left[\left\|V_{n,t}^{1/2}\zeta_{n,t}\right\|^3\right] &= \mathbb{E}\left[\left(\zeta_{n,t}^\top\mathbb{E}[X_{n,t}X_{n,t}^\top|\mathcal{F}_{n,t-1}]\zeta_{n,t}\right)^{3/2}\right] \\
&= \mathbb{E}\left[\int\left(u^\top\mathbb{E}[X_{n,t}X_{n,t}^\top|\mathcal{F}_{n,t-1}]u\right)^{3/2}\Phi\,(du)\right] \\
&= \mathbb{E}\left[\int\left(\mathbb{E}[(u^\top X_{n,t})^2|\mathcal{F}_{n,t-1}]\right)^{3/2}\Phi\,(du)\right] \\
&\leq \mathbb{E}\left[\int\mathbb{E}[|u^\top X_{n,t}|^3|\mathcal{F}_{n,t-1}]\Phi\,(du)\right] \\
&= \mathbb{E}\left[\left|\zeta_{n,t}^\top X_{n,t}\right|^3\right] = \sqrt{8/\pi}\,\mathbb{E}\left[\|X_{n,t}\|^3\right].
\end{aligned}$$

Note that $\Sigma_n - \xi_n$ is positive semi-definite. Hence,

$$\begin{aligned}\mathbb{E}\left[\left\|\xi_n^{1/2}\mathcal{N}\right\|^3\right] &= \mathbb{E}\left[\left(\mathcal{N}^\top\xi_n\mathcal{N}\right)^{3/2}\right] \le \mathbb{E}\left[\left(\mathcal{N}^\top\Sigma_n\mathcal{N}\right)^{3/2}\right] \\ &\le \lambda_{\max}(\Sigma_n)^{3/2}\mathbb{E}\left[\left(\mathcal{N}^\top\mathcal{N}\right)^{3/2}\right] \\ &\le \bar{\lambda}^{3/2}\left(\mathbb{E}\left[\left(\mathcal{N}^\top\mathcal{N}\right)^2\right]\right)^{3/4} \le 3\bar{\lambda}^{3/2}m_n^{3/2}.\end{aligned}$$

Hence, under the condition $K_n \ge 36\bar{\lambda}^3 m_n^3/B_n^2$,

$$\left|\mathbb{E}\left[f(S_n^*)\right] - \int f_n(x)F_n(dx)\right| \le \frac{3B_n}{\sigma_n^2\delta_n} + \frac{6\bar{\lambda}^{3/2}m_n^{3/2}}{\sigma_n^2\delta_n K_n^{1/2}} \le \frac{4B_n}{\sigma_n^2\delta_n}. \tag{A.21}$$

From (A.14) and (A.21),

$$\begin{aligned}\mathbb{P}\left(S_n^* \in A\right) &\le \frac{1}{1-\psi_n}\mathbb{E}\left[f_n\left(S_n^*\right)\right] \\ &\le \frac{1}{1-\psi_n}\left(\int f_n(x)F_n(dx) + \frac{4B_n}{\sigma_n^2\delta_n}\right) \\ &\le \frac{1}{1-\psi_n}\left(\psi_n + (1-\psi_n)F_n\left(A^{3\delta_n}\right) + \frac{4B_n}{\sigma_n^2\delta_n}\right) \\ &= F_n\left(A^{3\delta_n}\right) + \frac{1}{1-\psi_n}\left(\psi_n + \frac{4B_n}{\sigma_n^2\delta_n}\right),\end{aligned}$$

which proves (A.12) as wanted. *Q.E.D.*

Proof of Theorem 1. Let K_n satisfy the condition in Lemma A2 and then define S_n^* as in (A.3). The assertion of Theorem 1 then readily follows from Lemma A1 and Lemma A2. *Q.E.D.*

S.A.3 Proof of Corollary 1

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

$$\begin{aligned}\sum_{t=1}^{k_n}\mathbb{E}\left[\|X_{n,t}\|^3\right] &\le \sum_{t=1}^{k_n}\left(\mathbb{E}\left[\|X_{n,t}\|^4\right]\right)^{3/4} \\ &= \sum_{t=1}^{k_n}\left(\sum_{j=1}^{m_n}\sum_{l=1}^{m_n}\mathbb{E}\left[\left(X_{n,t}^{(j)}\right)^2\left(X_{n,t}^{(l)}\right)^2\right]\right)^{3/4} \\ &\le \sum_{t=1}^{k_n}\left(\sum_{j=1}^{m_n}\left(\mathbb{E}\left[\left(X_{n,t}^{(j)}\right)^4\right]\right)^{1/2}\right)^{3/2} \\ &= O\left(k_n^{-1/2}m_n^{3/2}\right),\end{aligned}$$

where the first inequality is by Jensen's inequality; the second inequality is by the Cauchy–Schwarz inequality and the last line follows from $\sup_{t,j} \mathbb{E}[(X_{n,t}^{(j)})^4] = O\left(k_n^{-2}\right)$. Plugging the estimate above into (3.4), we readily deduce the assertion of Corollary 1. *Q.E.D.*

S.A.4 Proof of Theorem 2

We first establish the martingale approximation as claimed in (3.6) and (3.7) in the main text; see Lemma A3 below. The variables $X_{n,t}^*$ and $\widetilde{X}_{n,t}$ are defined as follows:

$$X_{n,t}^* \equiv \sum_{s=-\infty}^{\infty} \left\{ \mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t}\right] - \mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t-1}\right] \right\}, \tag{A.22}$$

$$\widetilde{X}_{n,t} \equiv \sum_{s=0}^{\infty} \mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t-1}\right] - \sum_{s=0}^{\infty} \left\{ X_{t-s-1} - \mathbb{E}\left[X_{t-s-1}|\mathcal{F}_{n,t-1}\right] \right\}. \tag{A.23}$$

Lemma A3. *The following statements hold under Assumption 2 for each $j \in \{1, \ldots, m_n\}$*

(a) $\sum_{s=-\infty}^{\infty} \left\| \mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}\right] - \mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}\right] \right\|_q \leq 4\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s$;

(b) $\sup_{j,t,n} \|X_{n,t}^{*(j)}\|_q \leq 4\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s$ *and* $\mathbb{E}\left[X_{n,t}^*|\mathcal{F}_{n,t-1}\right] = 0$;

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

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

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

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

Proof of Lemma A3. (a) We first note that

$$\begin{aligned}
&\sum_{s=0}^{\infty} \left\| \mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}\right] - \mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}\right] \right\|_q \\
&\quad \leq \sum_{s=0}^{\infty} \left\| \mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}\right] \right\|_q + \sum_{s=0}^{\infty} \left\| \mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}\right] \right\|_q \\
&\quad \leq \bar{c}_n k_n^{-1/2} \left(\sum_{s=0}^{\infty} \psi_s + \sum_{s=0}^{\infty} \psi_{s+1} \right) \\
&\quad \leq 2\bar{c}_n k_n^{-1/2} \sum_{s=0}^{\infty} \psi_s < \infty.
\end{aligned} \tag{A.24}$$

In addition, we have

$$
\begin{aligned}
&\sum_{s=-\infty}^{-1}\left\|\mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}\right]-\mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}\right]\right\|_q \\
&\quad\le \sum_{s=-\infty}^{-1}\left\|X_{n,t+s}^{(j)}-\mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}\right]\right\|_q+\sum_{s=-\infty}^{-1}\left\|X_{n,t+s}^{(j)}-\mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}\right]\right\|_q \\
&\quad\le \sum_{s=1}^{\infty}\left\|X_{n,t-s}^{(j)}-\mathbb{E}\left[X_{n,t-s}^{(j)}|\mathcal{F}_{n,t}\right]\right\|_q+\sum_{s=1}^{\infty}\left\|X_{n,t-s}^{(j)}-\mathbb{E}\left[X_{n,t-s}^{(j)}|\mathcal{F}_{n,t-1}\right]\right\|_q \qquad (A.25)\\
&\quad\le \bar{c}_n k_n^{-1/2}\left(\sum_{s=1}^{\infty}\psi_{s+1}+\sum_{s=1}^{\infty}\psi_{s-1}\right) \\
&\quad\le 2\bar{c}_n k_n^{-1/2}\sum_{s=0}^{\infty}\psi_s<\infty.
\end{aligned}
$$

The assertion of part (a) then follows from (A.24) and (A.25).

(b) From (A.24) and (A.25), we deduce that $\|X_{n,t}^{*(j)}\|_q \le 4\bar{c}_n k_n^{-1/2}\sum_{s=0}^{\infty}\psi_s$. It remains to verify that $\mathbb{E}\left[X_{n,t}^{*}|\mathcal{F}_{n,t-1}\right]=0$. To this end, we set

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

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

$$
\left|X_{n,t}^{*(j)}(m)\right|\le\sum_{s=-\infty}^{\infty}\left|\mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t}\right]-\mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}\right]\right|
$$

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

(c) The assertion of part (c) follows from (3.5) directly. Indeed,

$$
\sum_{s=0}^{\infty}\left\|\mathbb{E}\left[X_{n,t+s}^{(j)}|\mathcal{F}_{n,t-1}\right]\right\|_q\le\bar{c}_n k_n^{-1/2}\sum_{s=0}^{\infty}\psi_s<\infty \qquad (A.26)
$$

and

$$
\sum_{s=0}^{\infty}\left\|X_{t-s-1}-\mathbb{E}\left[X_{t-s-1}|\mathcal{F}_{n,t-1}\right]\right\|_q\le\bar{c}_n k_n^{-1/2}\sum_{s=0}^{\infty}\psi_s<\infty. \qquad (A.27)
$$

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

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

$$
\begin{aligned}
&\widetilde{X}_{n,t+1}-\widetilde{X}_{n,t}+X_{n,t}\\
&\quad=\sum_{s=0}^{\infty}\mathbb{E}\left[X_{n,t+1+s}|\mathcal{F}_{n,t}\right]-\sum_{s=0}^{\infty}\left\{X_{t-s}-\mathbb{E}\left[X_{t-s}|\mathcal{F}_{n,t}\right]\right\}\\
&\qquad-\sum_{s=0}^{\infty}\mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t-1}\right]+\sum_{s=0}^{\infty}\left\{X_{t-s-1}-\mathbb{E}\left[X_{t-s-1}|\mathcal{F}_{n,t-1}\right]\right\}+X_{n,t}\\
&\quad=\sum_{s=1}^{\infty}\mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t}\right]+\sum_{s=0}^{\infty}\left\{\mathbb{E}\left[X_{t-s}|\mathcal{F}_{n,t}\right]-X_{t-s}\right\}\\
&\qquad-\sum_{s=0}^{\infty}\mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t-1}\right]+\sum_{s=1}^{\infty}\left\{X_{t-s}-\mathbb{E}\left[X_{t-s}|\mathcal{F}_{n,t-1}\right]\right\}+X_{n,t}\\
&\quad=\sum_{s=0}^{\infty}\left(\mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t}\right]-\mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t-1}\right]\right)+\sum_{s=1}^{\infty}\left\{\mathbb{E}\left[X_{t-s}|\mathcal{F}_{n,t}\right]-\mathbb{E}\left[X_{t-s}|\mathcal{F}_{n,t-1}\right]\right\}\\
&\quad=\sum_{s=-\infty}^{\infty}\left\{\mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t}\right]-\mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t-1}\right]\right\}=X_{n,t}^{*}.
\end{aligned}
$$

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

Proof of Theorem 2. By Theorem 1, Lemma A3(f) and the triangle inequality, there exists a sequence $\widetilde{S}_n^*$ of m_n-dimensional random vectors with distribution $\mathcal{N}(0, \Sigma_n^*)$ such that

$$\left\|S_n-\widetilde{S}_n^*\right\|=O_p(m_n^{1/2}r_n^{1/2}+(B_n^*m_n)^{1/3}+\bar{c}_n m_n^{1/2}k_n^{-1/2}), \tag{A.28}$$

where $\Sigma_n^* = \mathbb{E}[S_n^* S_n^{*\top}]$. By Lyapunov's inequality and Lemma A3(d),

$$\|S_n-S_n^*\|_2^2=\sum_{j=1}^{m_n}\mathbb{E}\left[\left|\widetilde{X}_{n,t}^{(j)}-\widetilde{X}_{n,k_n+1}^{(j)}\right|^2\right]\leq K\bar{c}_n^2 m_n k_n^{-1}. \tag{A.29}$$

By definition, $\Sigma_n - \Sigma_n^* = \mathbb{E}[S_n S_n^\top - S_n^* S_n^{*\top}]$. Hence, for any $a \in \mathbb{R}^{m_n}$,

$$
\begin{aligned}
\left\|a^\top\left(\Sigma_n-\Sigma_n^*\right)\right\|^2 \;\leq\; & K\left\|\mathbb{E}\left[a^\top(S_n-S_n^*)S_n^\top\right]\right\|^2+K\left\|\mathbb{E}\left[a^\top S_n(S_n-S_n^*)^\top\right]\right\|^2\\
&+K\left\|\mathbb{E}\left[a^\top(S_n-S_n^*)(S_n-S_n^*)^\top\right]\right\|^2.
\end{aligned} \tag{A.30}
$$

We now bound the terms on the majorant side of (A.30). Note that

$$\mathbb{E}\left[a^\top(S_n-S_n^*)S_n^\top\right]\Sigma_n^{-1}\mathbb{E}\left[S_n(S_n-S_n^*)^\top a\right]\leq\mathbb{E}\left[\left|a^\top(S_n-S_n^*)\right|^2\right],$$

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.30) can thus be bounded by

$$
\begin{aligned}
\left\|\mathbb{E}\left[a^\top(S_n-S_n^*)S_n^\top\right]\right\|^2 &\leq \lambda_{\max}(\Sigma_n)\mathbb{E}\left[\left|a^\top(S_n-S_n^*)\right|^2\right]\\
&\leq K\|a\|^2\|S_n-S_n^*\|_2^2
\end{aligned} \tag{A.31}
$$

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.30), we use the Cauchy–Schwarz inequality to derive

$$\left\|\mathbb{E}\left[a^\top S_n(S_n - S_n^*)^\top\right]\right\|^2 \leq \mathbb{E}\left[(a^\top S_n)^2\right]\|S_n - S_n^*\|_2^2 \leq K\|a\|^2\|S_n - S_n^*\|_2^2. \tag{A.32}$$

For the third term in (A.30), we observe

$$\begin{aligned}\left\|\mathbb{E}\left[a^\top(S_n - S_n^*)(S_n - S_n^*)^\top\right]\right\|^2 &\leq \|a\|^2\left(\lambda_{\max}\left(\mathbb{E}\left[(S_n - S_n^*)(S_n - S_n^*)^\top\right]\right)\right)^2 \\ &\leq \|a\|^2\left(\mathrm{Tr}\left(\mathbb{E}\left[(S_n - S_n^*)(S_n - S_n^*)^\top\right]\right)\right)^2 \\ &= \|a\|^2\|S_n - S_n^*\|_2^4. \end{aligned}\tag{A.33}$$

Combining (A.30)–(A.33), we deduce that

$$\begin{aligned}\sup_{\{a:\|a\|=1\}} a^\top(\Sigma_n - \Sigma_n^*)(\Sigma_n - \Sigma_n^*)^\top a &\leq K\|S_n - S_n^*\|_2^2 + K\|S_n - S_n^*\|_2^4 \\ &= O_p(\bar{c}_n^2 m_n k_n^{-1} + \bar{c}_n^4 m_n^2 k_n^{-2}).\end{aligned}$$

Hence,

$$\|\Sigma_n - \Sigma_n^*\|_S = O_p(\bar{c}_n m_n^{1/2} k_n^{-1/2} + \bar{c}_n^2 m_n k_n^{-1}). \tag{A.34}$$

Let $\widetilde{S}_n \equiv (\Sigma_n)^{1/2}(\Sigma_n^*)^{-1/2}\widetilde{S}_n^*$, so $\widetilde{S}_n \sim \mathcal{N}(0, \Sigma_n)$. By definition,

$$\widetilde{S}_n - \widetilde{S}_n^* = \left[(\Sigma_n)^{1/2} - (\Sigma_n^*)^{1/2}\right](\Sigma_n^*)^{-1/2}\widetilde{S}_n^*$$

which implies that

$$\begin{aligned}\mathbb{E}\left[\left\|\widetilde{S}_n - \widetilde{S}_n^*\right\|^2\right] &\leq \left\|(\Sigma_n)^{1/2} - (\Sigma_n^*)^{1/2}\right\|_S^2 \mathbb{E}\left[\widetilde{S}_n^{*\top}(\Sigma_n^*)^{-1}\widetilde{S}_n^*\right] \\ &\leq K\|\Sigma_n - \Sigma_n^*\|_S^2\,\mathbb{E}\left[\widetilde{S}_n^{*\top}(\Sigma_n^*)^{-1}\widetilde{S}_n^*\right] \\ &= O(\bar{c}_n^2 m_n^2 k_n^{-1} + \bar{c}_n^4 m_n^3 k_n^{-2})\end{aligned}\tag{A.35}$$

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 $\lambda_{\min}(\Sigma_n^*)^{-1} = O(1)$, and the last line follows from $\mathbb{E}[\widetilde{S}_n^{*\top}(\Sigma_n^*)^{-1}\widetilde{S}_n^*] = m_n$ and (A.34). Hence,

$$\left\|\widetilde{S}_n - \widetilde{S}_n^*\right\| = O_p(\bar{c}_n m_n k_n^{-1/2} + \bar{c}_n^2 m_n^{3/2} k_n^{-1}). \tag{A.36}$$

The assertion of the theorem then follows from (A.28) and (A.36). *Q.E.D.*

S.A.5 Proof of Theorem 3 and Theorem 4

Lemma A4. *Let $\Gamma_{X,n}^{(k,l)}(s) \equiv \mathbb{E}[X_{n,t}^{(k)} X_{n,t+s}^{(l)}]$. Under Assumption 2 and Assumption 4(i,iv),*

$$\max_{1\leq k,l\leq m_n}\sum_{s=-\infty}^{\infty}|s|^{r_2}|\Gamma_{X,n}^{(k,l)}(s)| \leq K\bar{c}_n^2 k_n^{-1}. \tag{A.37}$$

Proof of Lemma A4. For each $s \geq 0$,

$$\begin{aligned}
\left|\Gamma_{X,n}^{(k,l)}(s)\right| &= \left|\mathbb{E}[X_{n,t}^{(k)}\mathbb{E}[X_{n,t+s}^{(l)}|\mathcal{F}_{n,t}]]\right| \\
&\leq \left\|X_{n,t}^{(k)}\right\|_2 \left\|\mathbb{E}[X_{n,t+s}^{(l)}|\mathcal{F}_{n,t}]\right\|_2 \\
&\leq \left\|\mathbb{E}[X_{n,t}^{(k)}|\mathcal{F}_{n,t}]\right\|_q \left\|\mathbb{E}[X_{n,t+s}^{(l)}|\mathcal{F}_{n,t}]\right\|_q \\
&\leq \psi_0\psi_s\bar{c}_n^2 k_n^{-1}, \qquad (A.38)
\end{aligned}$$

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 2. Hence,

$$\sum_{s=-\infty}^{\infty} |s|^{r_2} \left|\Gamma_{X,n}^{(k,l)}(s)\right| \leq 2\sum_{s=0}^{\infty} s^{r_2} \left|\Gamma_{X,n}^{(k,l)}(s)\right| \leq \left(2\psi_0 \sum_{s=0}^{\infty} |s|^{r_2}\psi_s\right)\bar{c}_n^2 k_n^{-1}. \qquad (A.39)$$

By Assumption 4(iv), $K = 2\psi_0 \sum_{s=0}^{\infty} |s|^{r_2}\psi_s$ is finite. This finishes the proof. *Q.E.D.*

Lemma A5. *Under Assumptions 2, 3 and 4, we have for any s with $|s| \leq k_n - 1$,*

$$\max_{1\leq k,l\leq m_n} \left\|\widetilde{\Gamma}_{X,n}^{(k,l)}(s) - \mathbb{E}[\widetilde{\Gamma}_{X,n}^{(k,l)}(s)]\right\|_2^2 \leq K\bar{c}_n^4 k_n^{-1} \qquad (A.40)$$

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

Proof of Lemma A5. *Step 1.* In this step, we derive some preliminary estimates. Let $\eta_{t,s} = X_{n,t}^{(l)}X_{n,t+s}^{(k)} - \mathbb{E}[X_{n,t}^{(l)}X_{n,t+s}^{(k)}]$. We shall show that

$$|\mathbb{E}[\eta_{t,s}\eta_{t+h,s}]| \leq \begin{cases} \psi_{h-s}\bar{c}_n^4 k_n^{-2} & \text{when } h \geq s \geq 0, \\ K\left(\psi_{s-h} + \psi_h^2 + \psi_s^2\right)\bar{c}_n^4 k_n^{-2} & \text{when } s > h \geq 0. \end{cases} \qquad (A.41)$$

We start with the case $h \geq s \geq 0$. By Assumption 4(iii), we have for all $s \geq 0$,

$$\sup_t \max_{1\leq l,k\leq m_n} \mathbb{E}\left[\eta_{t,s}^2\right] \leq \sup_t \max_{1\leq l,k\leq m_n} \mathbb{E}\left[\left|X_{n,t}^{(l)}X_{n,t+s}^{(k)}\right|^2\right] \leq \bar{c}_n^4 k_n^{-2}. \qquad (A.42)$$

By the Cauchy–Schwarz inequality, Assumptions 4(ii) and (A.42), we deduce

$$|\mathbb{E}[\eta_{t,s}\eta_{t+h,s}]| = |\mathbb{E}[\eta_{t,s}\mathbb{E}[\eta_{t+h,s}|\mathcal{F}_{n,t+s}]]| \leq \|\eta_{t,s}\|_2 \|\mathbb{E}[\eta_{t+h,s}|\mathcal{F}_{n,t+s}]\|_2 \leq \psi_{h-s}\bar{c}_n^4 k_n^{-2}, \qquad (A.43)$$

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,

$$|\mathbb{E}[\eta_{t,s}\eta_{t+h,s}]| \leq |\mathbb{E}[\eta_{t,h}\eta_{t+s,h}]| + \left|\Gamma_{n,X}^{(k,l)}(h)\right|^2 + \left|\Gamma_{n,X}^{(k,l)}(s)\right|^2. \qquad (A.44)$$

By swapping s and h in (A.43), we obtain $|\mathbb{E}[\eta_{t,h}\eta_{t+s,h}]| \leq \psi_{s-h}\bar{c}_n^4 k_n^{-2}$. By (A.38),

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

The second claim in (A.41) then readily follows from these estimates.

Step 2. We now prove (A.40). Since $\widetilde{\Gamma}_{X,n}^{(k,l)}(s) = \widetilde{\Gamma}_{X,n}^{(l,k)}(-s)$, it suffices to consider $s \geq 0$. With $\eta_{t,s}$ defined in step 1, we can rewrite $\widetilde{\Gamma}_{X,n}^{(k,l)}(s) - \mathbb{E}[\widetilde{\Gamma}_{X,n}^{(k,l)}(s)] = \sum_{t=1}^{k_n-s}\eta_{t,s}$. Hence,

$$\begin{aligned}
\left\|\widetilde{\Gamma}_{X,n}^{(k,l)}(s) - \mathbb{E}[\widetilde{\Gamma}_{X,n}^{(k,l)}(s)]\right\|_2^2 &= \mathbb{E}\left[\left(\sum_{t=1}^{k_n-s}\eta_{t,s}\right)^2\right] \\
&\leq 2\sum_{h=0}^{k_n-s-1}\sum_{t=1}^{k_n-s-h}|\mathbb{E}[\eta_{t,s}\eta_{t+h,s}]| = 2\left(R_{1,n} + R_{2,n}\right), \qquad \text{(A.45)}
\end{aligned}$$

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

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

By (A.41),

$$R_{1,n} \leq \left(\sum_{h=s}^{k_n-s-1}\frac{k_n-s-h}{k_n}\psi_{h-s}\right)\bar{c}_n^4 k_n^{-1} \leq \left(\sum_{h=0}^{\infty}\psi_h\right)\bar{c}_n^4 k_n^{-1}, \qquad \text{(A.46)}$$

and similarly,

$$\begin{aligned}
R_{2,n} &\leq K\left(\sum_{h=0}^{(k_n-s-1)\wedge(s-1)}\frac{k_n-s-h}{k_n}\left(\psi_{s-h} + \psi_h^2 + \psi_s^2\right)\right)\bar{c}_n^4 k_n^{-1} \\
&\leq K\left(\sum_{h=0}^{\infty}\left(\psi_h + \psi_h^2\right) + s\psi_s^2\right)\bar{c}_n^4 k_n^{-1}. \qquad \text{(A.47)}
\end{aligned}$$

Combining (A.45), (A.46) and (A.47), we deduce

$$\left\|\widetilde{\Gamma}_{X,n}^{(k,l)}(s) - \mathbb{E}[\widetilde{\Gamma}_{X,n}^{(k,l)}(s)]\right\|_2^2 \leq K\left(\sum_{h=0}^{\infty}\psi_h + s\psi_s^2\right)\bar{c}_n^4 k_n^{-1}.$$

Since $\sum_{h=0}^{\infty}\psi_h < \infty$ and $\sup_{s\geq 0} s\psi_s^2 < \infty$ under Assumption 2 and Assumption 4(iv), the assertion of the lemma follows from the above inequality. *Q.E.D.*

Proof of Theorem 3. Recall that $\Gamma_{X,n}(s) \equiv \mathbb{E}[X_{n,t}X_{n,t+s}^\top]$. By definition, we can decompose

$$\begin{aligned}
\widetilde{\Sigma}_n - \Sigma_n &= \sum_{s=-k_n+1}^{k_n-1}\mathcal{K}(s/M_n)\left(\widetilde{\Gamma}_{X,n}(s) - \mathbb{E}\left[\widetilde{\Gamma}_{X,n}(s)\right]\right) \\
&\quad + \sum_{s=-k_n+1}^{k_n-1}(\mathcal{K}(s/M_n) - 1)(k_n - s)\Gamma_{X,n}(s). \qquad \text{(A.48)}
\end{aligned}$$

To bound the first term on the right-hand side of (A.48), we note, by the triangle inequality,

$$\begin{aligned}&\left\|\sum_{s=-k_n+1}^{k_n-1}\mathcal{K}\left(s/M_n\right)\left(\widetilde{\Gamma}_{X,n}(s)-\mathbb{E}\left[\widetilde{\Gamma}_{X,n}(s)\right]\right)\right\|\\&\qquad\leq\sum_{s=-k_n+1}^{k_n-1}\left|\mathcal{K}\left(s/M_n\right)\right|\left\|\widetilde{\Gamma}_{X,n}(s)-\mathbb{E}\left[\widetilde{\Gamma}_{X,n}(s)\right]\right\|.\end{aligned}\tag{A.49}$$

By (A.40),

$$\mathbb{E}\left[\left\|\widetilde{\Gamma}_{X,n}(s)-\mathbb{E}\left[\widetilde{\Gamma}_{X,n}(s)\right]\right\|\right]\leq\left(\sum_{k=1}^{m_n}\sum_{l=1}^{m_n}\left\|\widetilde{\Gamma}_{X,n}^{(k,l)}(s)-\mathbb{E}[\widetilde{\Gamma}_{X,n}^{(k,l)}(s)]\right\|_2^2\right)^{1/2}\leq K\bar{c}_n^2 m_n k_n^{-1/2}.$$

Combining this estimate with (A.49), we deduce

$$\begin{aligned}&\mathbb{E}\left[\left\|\sum_{s=-k_n+1}^{k_n-1}\mathcal{K}\left(s/M_n\right)\left(\widetilde{\Gamma}_{X,n}(s)-\mathbb{E}\left[\widetilde{\Gamma}_{X,n}(s)\right]\right)\right\|\right]\\&\qquad\leq K\bar{c}_n^2 m_n k_n^{-1/2}\sum_{s=-k_n+1}^{k_n-1}\left|\mathcal{K}\left(s/M_n\right)\right|\leq K\bar{c}_n^2 m_n M_n k_n^{-1/2},\end{aligned}$$

where the second inequality follows from Assumption 3. From here, we deduce

$$\sum_{s=-k_n+1}^{k_n-1}\mathcal{K}\left(s/M_n\right)\left(\widetilde{\Gamma}_{X,n}(s)-\mathbb{E}\left[\widetilde{\Gamma}_{X,n}(s)\right]\right)=O_p(\bar{c}_n^2 m_n M_n k_n^{-1/2}).\tag{A.50}$$

We now turn to the second term on the right-hand side of (A.48). By definition,

$$\begin{aligned}&\left\|\sum_{s=-k_n+1}^{k_n-1}(\mathcal{K}\left(s/M_n\right)-1)(k_n-s)\Gamma_{X,n}(s)\right\|^2\\&\qquad=\sum_{k=1}^{m_n}\sum_{l=1}^{m_n}\left|\sum_{s=-k_n+1}^{k_n-1}(\mathcal{K}\left(s/M_n\right)-1)(k_n-s)\Gamma_{X,n}^{(k,l)}(s)\right|^2.\end{aligned}\tag{A.51}$$

Let $r=r_1\wedge r_2$. By Assumption 3, we can fix some (small) constant $\varepsilon\in(0,1)$ such that

$$\frac{|1-\mathcal{K}(x)|}{|x|^r}\leq\frac{|1-\mathcal{K}(x)|}{|x|^{r_1}}\leq K\text{ for }x\in[-\varepsilon,\varepsilon].\tag{A.52}$$

By the triangle inequality

$$\begin{aligned}&\left|\sum_{s=-k_n+1}^{k_n-1}(\mathcal{K}\left(s/M_n\right)-1)(k_n-s)\Gamma_{X,n}^{(k,l)}(s)\right|\\&\qquad\leq k_nM_n^{-r}\sum_{|s|\leq\varepsilon M_n}\left|\frac{\mathcal{K}\left(s/M_n\right)-1}{|s/M_n|^r}\right||s|^r\left|\Gamma_{X,n}^{(k,l)}(s)\right|\\&\qquad\quad+k_n\sum_{\varepsilon M_n<|s|<k_n}\left|\mathcal{K}\left(s/M_n\right)-1\right|\left|\Gamma_{X,n}^{(k,l)}(s)\right|.\end{aligned}\tag{A.53}$$

By (A.52),

$$\sum_{|s|\leq\varepsilon M_n}\left|\frac{\mathcal{K}(s/M_n)-1}{|s/M_n|^r}\right||s|^r\left|\Gamma^{(k,l)}_{X,n}(s)\right| \leq K\sum_{|s|\leq\varepsilon M_n}|s|^r\left|\Gamma^{(k,l)}_{X,n}(s)\right| \leq K\sum_{s=-\infty}^{\infty}|s|^r\left|\Gamma^{(k,l)}_{X,n}(s)\right|. \tag{A.54}$$

Since $\mathcal{K}(\cdot)$ is bounded (Assumption 3),

$$\sum_{\varepsilon M_n<|s|<k_n}|\mathcal{K}(s/M_n)-1|\left|\Gamma^{(k,l)}_{X,n}(s)\right| \leq K\sum_{\varepsilon M_n<|s|<k_n}\left|\Gamma^{(k,l)}_{X,n}(s)\right| \leq KM_n^{-r}\sum_{s=-\infty}^{\infty}|s|^r\left|\Gamma^{(k,l)}_{X,n}(s)\right|. \tag{A.55}$$

Combining (A.53), (A.54) and (A.55), we deduce

$$\left|\sum_{s=-k_n+1}^{k_n-1}(\mathcal{K}(s/M_n)-1)(k_n-s)\Gamma^{(k,l)}_{X,n}(s)\right| \leq Kk_nM_n^{-r}\sum_{s=-\infty}^{\infty}|s|^r\left|\Gamma^{(k,l)}_{X,n}(s)\right| \leq K\bar{c}_n^2M_n^{-r} \tag{A.56}$$

where the second inequality is by (A.37). By (A.51) and (A.56),

$$\sum_{s=-k_n+1}^{k_n-1}(\mathcal{K}(s/M_n)-1)(k_n-s)\Gamma_{X,n}(s)=O(\bar{c}_n^2m_nM_n^{-r}). \tag{A.57}$$

The assertion of the theorem then follows from (A.48), (A.50) and (A.57). *Q.E.D.*

Proof of Theorem 4. By Theorem 3,

$$\left\|\widetilde{\Sigma}_n-\Sigma_n\right\|=O_p(\bar{c}_n^2m_n(M_nk_n^{-1/2}+M_n^{-r_1\wedge r_2})). \tag{A.58}$$

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

$$\left\|\widehat{\Sigma}_n-\widetilde{\Sigma}_n\right\|=O_p(M_nm_n^{1/2}\delta_{\theta,n}). \tag{A.59}$$

By the definitions of $\widehat{\Gamma}_{X,n}(s)$ and $\widetilde{\Gamma}_{X,n}(s)$, for any $s\geq 0$, we can decompose

$$\begin{aligned}\widehat{\Gamma}_{X,n}(s)-\widetilde{\Gamma}_{X,n}(s) &= k_n^{-1}\sum_{t=1}^{k_n-s}\left[g(Z_t,\hat{\theta}_n)-g(Z_t,\theta_0)\right]\left[g(Z_{t+s},\hat{\theta}_n)-g(Z_{t+s},\theta_0)\right]^\top \\ &\quad +k_n^{-1}\sum_{t=1}^{k_n-s}\left[g(Z_t,\hat{\theta}_n)-g(Z_t,\theta_0)\right]g(Z_{t+s},\theta_0)^\top \\ &\quad +k_n^{-1}\sum_{t=1}^{k_n-s}g(Z_t,\theta_0)\left[g(Z_{t+s},\hat{\theta}_n)-g(Z_{t+s},\theta_0)\right]^\top.\end{aligned} \tag{A.60}$$

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

$$\begin{aligned}
&\max_{|s|\leq k_n-1}\left\|\widehat{\Gamma}_{X,n}(s)-\widetilde{\Gamma}_{X,n}(s)\right\| \\
&\quad\leq k_n^{-1}\sum_{t=1}^{k_n}\left\|g(Z_t,\hat{\theta}_n)-g(Z_t,\theta_0)\right\|^2 \\
&\quad\quad+2\left(k_n^{-1}\sum_{t=1}^{k_n}\left\|g(Z_t,\hat{\theta}_n)-g(Z_t,\theta_0)\right\|^2\right)^{1/2}\left(k_n^{-1}\sum_{t=1}^{k_n}\|g(Z_t,\theta_0)\|^2\right)^{1/2}.
\end{aligned} \tag{A.61}$$

By Assumption 5(ii) and Markov's inequality,

$$k_n^{-1}\sum_{t=1}^{k_n}\|g(Z_t,\theta_0)\|^2=O_p(m_n). \tag{A.62}$$

By Assumption 5(i), (A.61) and (A.62), we deduce

$$\max_{|s|\leq k_n-1}\left\|\widehat{\Gamma}_{X,n}(s)-\widetilde{\Gamma}_{X,n}(s)\right\|=O_p(m_n^{1/2}\delta_{\theta,n}). \tag{A.63}$$

By the triangle inequality, (A.63) and Assumption 3(i), we deduce

$$\begin{aligned}
\left\|\widehat{\Sigma}_n-\widetilde{\Sigma}_n\right\| &\leq \sum_{s=-k_n+1}^{k_n-1}|\mathcal{K}(s/M_n)|\left\|\widehat{\Gamma}_{X,n}(s)-\widetilde{\Gamma}_{X,n}(s)\right\| \\
&\leq \max_{|s|\leq k_n-1}\left\|\widehat{\Gamma}_{X,n}(s)-\widetilde{\Gamma}_{X,n}(s)\right\|\sum_{s=-k_n+1}^{k_n-1}|\mathcal{K}(s/M_n)| \\
&= O_p(M_n m_n^{1/2}\delta_{\theta,n})
\end{aligned} \tag{A.64}$$

as claimed in (A.59). This finishes the proof. *Q.E.D.*

S.A.6 Proof of Theorem 5

Proof of Theorem 5. *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

$$\frac{n^{1/2}P(x)^\top(\hat{b}_n-b_n^*)}{\sigma_n(x)}=\frac{n^{-1/2}P(x)^\top Q_n^{-1}P_n^\top U_n}{\sigma_n(x)}+O_{pu}(\delta_{1,n}+m_n^{1/2}\delta_{3,n}), \tag{A.65}$$

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

By Assumption 6(ii),

$$\sup_{x\in\mathcal{X}}\frac{\|P(x)\|}{\sigma_n(x)}\leq(\lambda_{\min}(A_n))^{-1/2}\lambda_{\max}(Q_n)\leq K. \tag{A.66}$$

By Assumptions 6(ii), (iv) and (v), we have, with probability approaching one,

$$\lambda_{\max}(\widehat{Q}_n) + \lambda_{\max}(\widehat{A}_n) + \lambda_{\min}^{-1}(\widehat{Q}_n) + \lambda_{\min}^{-1}(\widehat{A}_n) \leq K. \tag{A.67}$$

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

$$\hat{b}_n - b_n^* = (P_n^\top P_n)^{-1} \left(P_n^\top U_n\right) + (P_n^\top P_n)^{-1} P_n^\top (H_n - H_n^*). \tag{A.68}$$

By Assumption 6(ii),

$$\mathbb{E}\left[\left\|n^{-1} P_n^\top U_n\right\|^2\right] = n^{-1} \mathrm{Tr}\,(A_n) \leq K m_n n^{-1}, \tag{A.69}$$

which together with Markov's inequality implies that

$$\left\|n^{-1/2} P_n^\top U_n\right\| = O_p(m_n^{1/2}). \tag{A.70}$$

We observe

$$\begin{aligned}
&\sup_{x\in\mathcal{X}} \frac{1}{\sigma_n(x)} \left|n^{1/2} P(x)^\top (P_n^\top P_n)^{-1} \left(P_n^\top U_n\right) - n^{-1/2} P(x)^\top Q_n^{-1} \left(P_n^\top U_n\right)\right| \\
&\quad = \sup_{x\in\mathcal{X}} \frac{1}{\sigma_n(x)} \left|P(x)^\top \widehat{Q}_n^{-1} \left(\widehat{Q}_n - Q_n\right) Q_n^{-1} \left(n^{-1/2} P_n^\top U_n\right)\right| \\
&\quad \leq (\lambda_{\min}(\widehat{Q}_n)\lambda_{\min}(Q_n))^{-1} \left\|\widehat{Q}_n - Q_n\right\| \left\|n^{-1/2} P_n^\top U_n\right\| \sup_{x\in\mathcal{X}} \frac{\|P(x)\|}{\sigma_n(x)} \\
&\quad = O_p(m_n^{1/2} \delta_{3,n}),
\end{aligned} \tag{A.71}$$

where the inequality follows from the Cauchy–Schwarz inequality and the last line is derived from Assumption 6(iv), (A.66), (A.67) and (A.70).

By Assumption 6(i), (A.66) and (A.67),

$$\begin{aligned}
&\sup_{x\in\mathcal{X}} \frac{1}{\sigma_n(x)} \left|n^{1/2} P(x)^\top (P_n^\top P_n)^{-1} P_n^\top (H_n - H_n^*)\right| \\
&\quad \leq \sup_{x\in\mathcal{X}} \frac{\|P(x)\|}{\sigma_n(x)} \left((H_n - H_n^*)^\top P_n (P_n^\top P_n)^{-1/2} \widehat{Q}_n^{-1} (P_n^\top P_n)^{-1/2} P_n^\top (H_n - H_n^*)\right)^{1/2} \\
&\quad \leq \frac{\left((H_n - H_n^*)^\top P_n (P_n^\top P_n)^{-1} P_n^\top (H_n - H_n^*)\right)^{1/2}}{(\lambda_{\min}(\widehat{Q}_n))^{1/2}} \sup_{x\in\mathcal{X}} \frac{\|P(x)\|}{\sigma_n(x)} \\
&\quad \leq \frac{\sup_{x\in\mathcal{X}} n^{1/2} \left|h(x) - P(x)^\top b_n^*\right|}{(\lambda_{\min}(\widehat{Q}_n))^{1/2}} \sup_{x\in\mathcal{X}} \frac{\|P(x)\|}{\sigma_n(x)} = O_p(\delta_{1,n}).
\end{aligned} \tag{A.72}$$

The claim in (A.65) follows by combining the results in (A.68), (A.71) and (A.72).

Step 2. In this step, we show that

$$\sup_{x\in\mathcal{X}} \left|\frac{n^{1/2} P(x)^\top (\hat{b}_n - b_n^*)}{\hat{\sigma}_n(x)} - \frac{n^{1/2} P(x)^\top (\hat{b}_n - b_n^*)}{\sigma_n(x)}\right| = O_p(m_n^{1/2}(\delta_{3,n} + \delta_{4,n})). \tag{A.73}$$

By the triangle inequality

$$\begin{aligned}\left\|\widehat{\Sigma}_n-\Sigma_n\right\|_S &\leq \left\|(\widehat{Q}_n^{-1}-Q_n^{-1})\widehat{A}_n\widehat{Q}_n^{-1}\right\|_S \\ &\quad +\left\|Q_n^{-1}(\widehat{A}_n-A_n)\widehat{Q}_n^{-1}\right\|_S+\left\|Q_n^{-1}A_n(\widehat{Q}_n^{-1}-Q_n^{-1})\right\|_S.\end{aligned}$$

By the Cauchy–Schwarz inequality, Assumption 6(ii, iv) and (A.67),

$$\left\|(\widehat{Q}_n^{-1}-Q_n^{-1})\widehat{A}_n\widehat{Q}_n^{-1}\right\|_S \leq \frac{\lambda_{\max}(\widehat{A}_n)\left\|\widehat{Q}_n-Q_n\right\|}{\lambda_{\min}(Q_n)(\lambda_{\min}(\widehat{Q}_n))^2}=O_p(\delta_{3,n}).$$

Similarly, $\|Q_n^{-1}(\widehat{A}_n-A_n)\widehat{Q}_n^{-1}\|_S=O_p(\delta_{4,n})$ and $\|Q_n^{-1}A_n(\widehat{Q}_n^{-1}-Q_n^{-1})\|_S=O_p(\delta_{3,n})$. Combining these estimates, we get

$$\left\|\widehat{\Sigma}_n-\Sigma_n\right\|_S=O_p(\delta_{3,n}+\delta_{4,n}). \tag{A.74}$$

By Assumption 6(ii), this estimate further implies that, with probability approaching one,

$$\lambda_{\min}^{-1}(\widehat{\Sigma}_n)\leq K, \quad \lambda_{\max}(\widehat{\Sigma}_n)\leq K. \tag{A.75}$$

We then observe

$$\begin{aligned}\sup_{x\in\mathcal{X}}\frac{|\sigma_n(x)-\hat{\sigma}_n(x)|}{\hat{\sigma}_n(x)} &= \sup_{x\in\mathcal{X}}\frac{\left|\sigma_n^2(x)-\hat{\sigma}_n^2(x)\right|}{\hat{\sigma}_n(x)(\sigma_n(x)+\hat{\sigma}_n(x))} \\ &= \sup_{x\in\mathcal{X}}\frac{\left|P(x)^\top(\widehat{\Sigma}_n-\Sigma_n)P(x)\right|}{\hat{\sigma}_n(x)(\sigma_n(x)+\hat{\sigma}_n(x))} \\ &\leq \frac{\left\|\widehat{\Sigma}_n-\Sigma_n\right\|_S}{(\lambda_{\min}(\widehat{\Sigma}_n)\lambda_{\min}(\Sigma_n))^{1/2}+\lambda_{\min}(\widehat{\Sigma}_n)} \\ &= O_p(\delta_{3,n}+\delta_{4,n})\end{aligned} \tag{A.76}$$

where the last line follows from Assumption 6(ii), (A.74) and (A.75).

By the Cauchy–Schwarz inequality

$$\sup_{x\in\mathcal{X}}\left|\frac{P(x)^\top Q_n^{-1}}{\sigma_n(x)}n^{-1/2}\sum_{t=1}^n u_tP(X_t)\right| \leq \frac{\left\|n^{-1/2}P_n^\top U_n\right\|}{\lambda_{\min}(Q_n)}\sup_{x\in\mathcal{X}}\frac{\|P(x)\|}{\sigma_n(x)}=O_p(m_n^{1/2}) \tag{A.77}$$

where the equality is due to Assumption 6(ii), (A.66) and (A.70). By (A.65) and (A.77),

$$\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}P(x)^\top(\hat{b}_n-b_n^*)}{\sigma_n(x)}\right|=O_p(m_n^{1/2}). \tag{A.78}$$

Combining (A.76) and (A.78), we deduce

$$\begin{aligned}&\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}P(x)^\top(\hat{b}_n-b_n^*)}{\hat{\sigma}_n(x)}-\frac{n^{1/2}P(x)^\top(\hat{b}_n-b_n^*)}{\sigma_n(x)}\right| \\ &\quad\leq \sup_{x\in\mathcal{X}}\frac{|\sigma_n(x)-\hat{\sigma}_n(x)|}{\hat{\sigma}_n(x)}\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}P(x)^\top(\hat{b}_n-b_n^*)}{\sigma_n(x)}\right|=O_p(m_n^{1/2}(\delta_{3,n}+\delta_{4,n})),\end{aligned}$$

which finishes the proof of (A.73).

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

$$\frac{n^{1/2}\left(\hat{h}_n(x)-h(x)\right)}{\hat{\sigma}_n(x)}=\frac{P(x)^\top \widetilde{S}_n}{\sigma_n(x)}+O_{pu}(\delta_n), \tag{A.79}$$

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

By (A.65) and (A.73),

$$\frac{n^{1/2}P(x)^\top(\hat{b}_n-b_n^*)}{\hat{\sigma}_n(x)}=\frac{n^{-1/2}P(x)^\top Q_n^{-1}P_n^\top U_n}{\sigma_n(x)}+O_{pu}(\delta_{1,n}+m_n^{1/2}(\delta_{3,n}+\delta_{4,n})). \tag{A.80}$$

By Assumption 6(ii) and (A.66),

$$\sup_{x\in\mathcal{X}}\frac{\left\|P(x)^\top Q_n^{-1}\right\|}{\sigma_n(x)}\leq K. \tag{A.81}$$

Hence, by Assumption 6(iii), (A.80) and (A.81),

$$\frac{n^{1/2}P(x)^\top(\hat{b}_n-b_n^*)}{\hat{\sigma}_n(x)}=\frac{P(x)^\top Q_n^{-1}\widetilde{N}_n}{\sigma_n(x)}+O_{pu}(\delta_{1,n}+\delta_{2,n}+m_n^{1/2}(\delta_{3,n}+\delta_{4,n})). \tag{A.82}$$

By Assumption 6(i) and (A.66)

$$\sup_{x\in\mathcal{X}}\frac{\left|n^{1/2}(h(x)-P(x)^\top b_n^*)\right|}{\sigma_n(x)}=\sup_{x\in\mathcal{X}}\frac{\|P(x)\|}{\sigma_n(x)}\frac{\left|n^{1/2}(h(x)-P(x)^\top b_n^*)\right|}{\|P(x)\|}=O(\delta_{1,n}), \tag{A.83}$$

which combined with (A.76) yields

$$\begin{aligned}\sup_{x\in\mathcal{X}}\frac{\left|n^{1/2}(h(x)-P(x)^\top b_n^*)\right|}{\hat{\sigma}_n(x)} &\leq \sup_{x\in\mathcal{X}}\frac{|\hat{\sigma}_n(x)-\sigma_n(x)|}{\hat{\sigma}_n(x)}\frac{\left|n^{1/2}(h(x)-P(x)^\top b_n^*)\right|}{\sigma_n(x)}\\ &\quad+\sup_{x\in\mathcal{X}}\frac{\left|n^{1/2}(h(x)-P(x)^\top b_n^*)\right|}{\sigma_n(x)}\\ &= O_p(\delta_{1,n}).\end{aligned} \tag{A.84}$$

From (A.82) and (A.84), the assertion (A.79) 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.7 Proof of Theorem 6

We first introduce some notations and a preliminary estimate; see Lemma A6, 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

$$\hat{b}_n^\dagger \equiv \left(\sum_{t=1}^n P\left(X_t\right) P\left(X_t\right)^\top\right)^{-1} \left(\sum_{t=1}^n P\left(X_t\right) g\left(Y_t^*; \gamma_0\right)\right).$$

Lemma A6. *Let* $\delta_{6,n} \equiv \sup_{x\in\mathcal{X}} \|P(x)\|^{-1}$*. Under Assumption 8,*

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

Proof of Lemma A6. *Step 1.* In this step, we show that

$$\sup_{x\in\mathcal{X}} \frac{n^{1/2}P(x)^\top(\hat{b}_n - \hat{b}_n^\dagger)}{\sigma_n(x)} = \sup_{x\in\mathcal{X}} \frac{n^{1/2}P(x)^\top Q_n^{-1} G_n(\hat{\gamma}_n - \gamma_0)}{\sigma_n(x)} + O_p(\delta_{3,n} + \delta_{5,n}). \tag{A.85}$$

By definition,

$$\hat{b}_n - \hat{b}_n^\dagger = \left(P_n^\top P_n\right)^{-1} \sum_{t=1}^n P(X_t)(g(Y_t^*, \hat{\gamma}_n) - g(Y_t^*, \gamma_0)). \tag{A.86}$$

Applying the second-order Taylor expansion, we further deduce

$$\begin{aligned} \hat{b}_n - \hat{b}_n^\dagger = &\left(P_n^\top P_n\right)^{-1} \sum_{t=1}^n P(X_t) g_\gamma(Y_t^*, \gamma_0)^\top (\hat{\gamma}_n - \gamma_0) \\ &+ \frac{1}{2}\left(P_n^\top P_n\right)^{-1} \sum_{t=1}^n P(X_t)(\hat{\gamma}_n - \gamma_0)^\top g_{\gamma\gamma}(Y_t^*, \widetilde{\gamma}_n)(\hat{\gamma}_n - \gamma_0), \end{aligned} \tag{A.87}$$

where $\widetilde{\gamma}_n$ is a mean value between $\hat{\gamma}_n$ and γ_0 that may vary across rows. By Assumption 8(iv,vi),

$$n^{-1} \sum_{t=1}^n \left|(\hat{\gamma}_n - \gamma_0)^\top g_{\gamma\gamma}(Y_t^*, \widetilde{\gamma}_n)(\hat{\gamma}_n - \gamma_0)\right|^2 = O_p(n^{-2}). \tag{A.88}$$

Since $P_n \left(P_n^\top P_n\right)^{-1} P_n^\top$ is a projection matrix,

$$\begin{aligned} &\left\|\left(P_n^\top P_n\right)^{-1} \sum_{t=1}^n P(X_t)(\hat{\gamma}_n - \gamma_0)^\top g_{\gamma\gamma}(Y_t^*, \widetilde{\gamma}_n)(\hat{\gamma}_n - \gamma_0)\right\|^2 \\ &\quad \leq \lambda_{\min}^{-1}(\widehat{Q}_n) n^{-1} \sum_{t=1}^n \left|(\hat{\gamma}_n - \gamma_0)^\top g_{\gamma\gamma}(Y_t^*, \widetilde{\gamma}_n)(\hat{\gamma}_n - \gamma_0)\right|^2 = O_p(n^{-2}), \end{aligned} \tag{A.89}$$

where the rate statement follows from (A.88) and Assumption 7. Collecting the results in (A.87) and (A.89), we get

$$\hat{b}_n - \hat{b}_n^\dagger = \left(P_n^\top P_n\right)^{-1} \sum_{t=1}^n P(X_t) g_\gamma(Y_t^*, \gamma_0)^\top (\hat{\gamma}_n - \gamma_0) + O_p(n^{-1}). \tag{A.90}$$

By Assumptions 7 and 8(ii,vi),

$$\left(P_n^\top P_n\right)^{-1}\sum_{t=1}^n P(X_t)g_\gamma(Y_t^*,\gamma_0)^\top(\hat\gamma_n-\gamma_0)=\widehat{Q}_n^{-1}G_n(\hat\gamma_n-\gamma_0)+O_p(\delta_{5,n}n^{-1/2}). \tag{A.91}$$

For $1\le j\le d$, let $g_{\gamma,j}(Y_t^*,\gamma_0)$ denote the jth component of $g_\gamma(Y_t^*,\gamma_0)$ and let $G_{n,j}$ denote jth column of G_n. We note that

$$G_{n,j}^\top Q_n^{-1}G_{n,j}\le n^{-1}\sum_{t=1}^n\mathbb{E}\left[g_{\gamma,j}(Y_t^*,\gamma_0)^2\right], \tag{A.92}$$

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 $P\left(X_t\right)$ under the product measure $\mathbb{P}\otimes P_n$, with P_n being the empirical measure. Hence, for any $1\le j\le d$,

$$\|G_{n,j}\|^2\le\frac{\lambda_{\max}(Q_n)}{n}\sum_{t=1}^n\mathbb{E}\left[g_{\gamma,j}(Y_t^*,\gamma_0)^2\right]\le\lambda_{\max}(Q_n)\sup_t\mathbb{E}\left[\|g_\gamma(Y_t^*,\gamma_0)\|^2\right]. \tag{A.93}$$

By Assumptions 6(ii) and 8(iv), we further deduce

$$\|G_n\|^2\le K. \tag{A.94}$$

By Assumption 7, Assumption 8(iii) and (A.94),

$$\begin{aligned}
&\left\|\left(\widehat{Q}_n^{-1}-Q_n^{-1}\right)G_n(\hat\gamma_n-\gamma_0)\right\|^2\\
&\quad=\left\|\widehat{Q}_n^{-1}\left(\widehat{Q}_n-Q_n\right)Q_n^{-1}G_n(\hat\gamma_n-\gamma_0)\right\|^2\\
&\quad\le\frac{\|\hat\gamma_n-\gamma_0\|^2}{(\lambda_{\min}(\widehat{Q}_n)\lambda_{\min}(Q_n))^2}\left\|\widehat{Q}_n-Q_n\right\|^2\|G_n\|^2=O_p(n^{-1}\delta_{3,n}^2)
\end{aligned}$$

which further implies that

$$\left(\widehat{Q}_n^{-1}-Q_n^{-1}\right)G_n(\hat\gamma_n-\gamma_0)=O_p(n^{-1/2}\delta_{3,n}). \tag{A.95}$$

Combining the results in (A.91) and (A.95), we get

$$\hat b_n-\hat b_n^\dagger=Q_n^{-1}G_n(\hat\gamma_n-\gamma_0)+O_p((\delta_{3,n}+\delta_{5,n})n^{-1/2}). \tag{A.96}$$

With an appeal to the Cauchy–Schwarz inequality, we deduce (A.85) from (A.96) and (A.66).

Step 2. We now prove the assertion of Lemma A6. For $j\in\{1,\ldots,d\}$, let $\phi_{n,j}^*$ denote the jth column of ϕ_n^*; recall the definition of ϕ_n^* from Assumption 8. By definition,

$$\begin{aligned}
&P(x)^\top Q_n^{-1}G_{n,j}-P(x)^\top\phi_{n,j}^*\\
&\quad=P(x)^\top Q_n^{-1}\left(G_{n,j}-Q_n\phi_{n,j}^*\right)\\
&\quad=n^{-1}\sum_{t=1}^n P(x)^\top Q_n^{-1}\left(\mathbb{E}\left[P(X_t)\left(H_j(X_t)-H_j^*(X_t)\right)\right]\right),
\end{aligned} \tag{A.97}$$

where $H_j(X_t) = \mathbb{E}\left[g_{\gamma,j}(Y_t^*, \gamma_0)\right| X_t]$ 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.92), we can show that

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

which together with Assumption 7 and Assumption 8(iii) implies that

$$\left\|\mathbb{E}\left[n^{-1}\sum_{t=1}^n P(X_t)\left(H_j(X_t) - H_j^*(X_t)\right)\right]\right\| = O(m_n^{-\rho}). \tag{A.98}$$

By (A.98) and the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\sup_{x\in\mathcal{X}}\left|n^{-1}\sum_{t=1}^n P(x)^\top Q_n^{-1}\mathbb{E}\left[P(X_t)\left(H_j(X_t) - H_j^*(X_t)\right)\right]\right|^2 \\
&\quad\leq \sup_{x\in\mathcal{X}}\|P(x)\|^2\left(\lambda_{\min}(Q_n)\right)^{-2}\left\|\mathbb{E}\left[n^{-1}\sum_{t=1}^n P(X_t)\left(H_j(X_t) - H_j^*(X_t)\right)\right]\right\|^2 \\
&\quad\leq K m_n^{1-2\rho}\zeta_n^2 = O(1)
\end{aligned}$$

where the last line is by (A.98), Assumption 7 and Assumption 8(v). By (A.97), we further deduce that

$$\sup_{x\in\mathcal{X}}\left\|P(x)^\top Q_n^{-1}G_n - P(x)^\top\phi_n^*\right\| \leq K. \tag{A.99}$$

From Assumption 8(iii,iv), it is easy to see that $P\left(\cdot\right)^\top \phi_n^*$ is uniformly bounded. Hence, by (A.99),

$$\sup_{x\in\mathcal{X}}\left\|P(x)^\top Q_n^{-1}G_n\right\| \leq K. \tag{A.100}$$

Using the Cauchy–Schwarz inequality, we deduce from (A.100), Assumption 7 and Assumption 8(vi) that

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

The assertion of Lemma A6 then follows from this estimate and (A.85). *Q.E.D.*

Proof of Theorem 6. By Assumption 7, we can invoke (A.76) in the proof of Theorem 5 to get

$$\sup_{x\in\mathcal{X}}\frac{\left|\sigma_n\left(x\right) - \hat{\sigma}_n\left(x\right)\right|}{\hat{\sigma}_n\left(x\right)} = O_p(\delta_{3,n} + \delta_{4,n}), \tag{A.101}$$

which together with Lemma A6 implies that

$$
\begin{aligned}
&\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}P(x)^\top(\hat{b}_n-\hat{b}_n^\dagger)}{\hat{\sigma}_n(x)}-\frac{n^{1/2}P(x)^\top(\hat{b}_n-\hat{b}_n^\dagger)}{\sigma_n(x)}\right| \\
&\quad\le \sup_{x\in\mathcal{X}}\frac{|\sigma_n(x)-\hat{\sigma}_n(x)|}{\hat{\sigma}_n(x)}\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}P(x)^\top(\hat{b}_n-\hat{b}_n^\dagger)}{\sigma_n(x)}\right| \qquad \text{(A.102)}\\
&\quad= O_p((\delta_{3,n}+\delta_{5,n}+\delta_{6,n})(\delta_{3,n}+\delta_{4,n})).
\end{aligned}
$$

Therefore,

$$
\sup_{x\in\mathcal{X}}\frac{n^{1/2}P(x)^\top(\hat{b}_n-\hat{b}_n^\dagger)}{\hat{\sigma}_n(x)}=O_p(\delta_{3,n}+\delta_{5,n}+\delta_{6,n}). \qquad \text{(A.103)}
$$

Let $\hat{h}_n^\dagger(x)=P(x)^\top\hat{b}_n^\dagger$. Applying (A.79) with $\hat{h}_n^\dagger(x)$ replacing $\hat{h}_n(x)$,

$$
\frac{n^{1/2}\left(\hat{h}_n^\dagger(x)-h(x)\right)}{\hat{\sigma}_n(x)}=\frac{P(x)^\top\widetilde{S}_n}{\sigma_n(x)}+O_{pu}(\delta_n).
$$

Then, by Lemma A6,

$$
\frac{n^{1/2}\left(\hat{h}_n(x)-h(x)\right)}{\hat{\sigma}_n(x)}=\frac{P(x)^\top\widetilde{S}_n}{\sigma_n(x)}+O_{pu}(\delta_n+\delta_{5,n}+\delta_{6,n}). \qquad \text{(A.104)}
$$

Under the null hypothesis (3.18) with $h(x)=0$,

$$
\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}\hat{h}_n(x)}{\hat{\sigma}_n(x)}\right|=\sup_{x\in\mathcal{X}}\left|\frac{P(x)^\top\widetilde{S}_n}{\sigma_n(x)}\right|+O_{pu}(\delta_n+\delta_{5,n}+\delta_{6,n}). \qquad \text{(A.105)}
$$

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 6 then follows from (A.105) 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.101) and (A.104),

$$
\begin{aligned}
\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}\hat{h}_n(x)}{\hat{\sigma}_n(x)}\right| &\ge \sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}h(x)}{\hat{\sigma}_n(x)}\right|-\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}\left(\hat{h}_n(x)-h(x)\right)}{\hat{\sigma}_n(x)}\right| \\
&= \sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}h(x)}{\sigma_n(x)}\frac{\sigma_n(x)}{\hat{\sigma}_n(x)}\right|-\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}\left(\hat{h}_n(x)-h(x)\right)}{\hat{\sigma}_n(x)}\right| \\
&\ge \sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}h(x)}{\sigma_n(x)}\right|\left(1-\sup_{x\in\mathcal{X}}\left|\frac{\sigma_n(x)}{\hat{\sigma}_n(x)}-1\right|\right)-\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}\left(\hat{h}_n(x)-h(x)\right)}{\hat{\sigma}_n(x)}\right| \\
&= \sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}h(x)}{\sigma_n(x)}\right|(1+o_p(1))-\sup_{x\in\mathcal{X}}\left|\frac{P(x)^\top\widetilde{S}_n}{\sigma_n(x)}\right|+o_p(1). \qquad \text{(A.106)}
\end{aligned}
$$

By the Cauchy–Schwarz inequality and Assumption 7,

$$\sup_{x\in\mathcal{X}}\left|\frac{P(x)^{\top}\widetilde{S}_n}{\sigma_n(x)}\right| \leq \frac{\lambda_{\max}(Q_n)\left\|\widetilde{S}_n\right\|}{(\lambda_{\min}(A_n))^{1/2}} = O_p(m_n^{1/2}). \tag{A.107}$$

Since $\mathbb{E}\left[g(Y_t^*,\gamma_0)|X_t = x\right] \neq 0$ 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 8(v), $\sup_{x\in\mathcal{X}}\|P(x)\| \leq \zeta_n m_n^{1/2}$. Therefore,

$$\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}h(x)}{\sigma_n(x)}\right| \geq \frac{n^{1/2}\lambda_{\min}(Q_n)c_0}{(\lambda_{\max}(A_n))^{1/2}\zeta_n m_n^{1/2}}, \tag{A.108}$$

which together with (A.106) and (A.107) implies that (recalling $\zeta_n m_n \ll n^{1/2}$ from Assumption 8(vii)),

$$\sup_{x\in\mathcal{X}}\left|\frac{n^{1/2}\hat{h}_n(x)}{\hat{\sigma}_n(x)}\right| \geq \frac{n^{1/2}\lambda_{\min}(Q_n)c_0}{(\lambda_{\max}(A_n))^{1/2}\zeta_n m_n^{1/2}}\left(1+o_p(1)\right). \tag{A.109}$$

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 7(ii) and 8(vii) implies that

$$\frac{(\lambda_{\max}(A_n))^{1/2}cv_{n,\alpha}\zeta_n m_n^{1/2}}{\lambda_{\min}(Q_n)(1+o_p(1))n^{1/2}} = o_p(1). \tag{A.110}$$

Combining the results in (A.109) and (A.110), we deduce that

$$\begin{aligned}\Pr\left(\widehat{T}_n \leq cv_{n,\alpha}\right) &\leq \Pr\left(\frac{n^{1/2}\lambda_{\min}(Q_n)c_0\left(1+o_p(1)\right)}{(\lambda_{\max}(A_n))^{1/2}\zeta_n m_n^{1/2}} \leq cv_{n,\alpha}\right)\\ &= \Pr\left(c_0 \leq \frac{(\lambda_{\max}(A_n))^{1/2}cv_{n,\alpha}\zeta_n m_n^{1/2}}{\lambda_{\min}(Q_n)(1+o_p(1))n^{1/2}}\right) \to 0\end{aligned}$$

as $n\to\infty$. From here, the second assertion readily follows. *Q.E.D.*

S.B Additional technical results

This section collects additional technical results that are mentioned in the main text. Section S.B.1 provides examples for the strong approximation result under primitive conditions. Section S.B.2 provides primitive conditions for Assumption 6 in the main text. Section S.B.3 provides details on how to verify the high-level conditions in Chernozhukov, Lee, and Rosen (2013) for the purpose of extending the inference for intersection bounds to time-series applications. S.B.4 provides details on how to derive the conditional moment restriction used in our empirical analysis.

S.B.1 Examples for Theorem 2 under primitive conditions

Condition (ii) of Theorem 2 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 $(\varepsilon_{n,t}, \mathcal{F}_{n,t})$ be a martingale difference array such that $\|\varepsilon_{n,t}^{(j)}\|_q \leq \bar{c}_n k_n^{-1/2}$ 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 (θ_j) satisfy $\sum_{|j|<\infty} |j\theta_j| < \infty$. Then $(X_{n,t})$ is an L_q-mixingale that satisfies Assumption 2 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 2 if and only if $\varepsilon_{n,t}$ satisfies Assumption 1. In the simple case when $\varepsilon_{n,t}$ has constant covariance matrix Σ_ε, 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 (3.8) is absent. $\square$

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)* $\sup_{t,j} ||\varepsilon_t^{(j)}||_\kappa \leq c_{\kappa,n}$ *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)* $\sum_{s=0}^{\infty} \varphi_s^{(\kappa-4)/(5\kappa)} < \infty$ *for the strong mixing case and* $\sum_{s=0}^{\infty} \varphi_s^{1/2} < \infty$ *for the uniform mixing case; (iii) the eigenvalues of* $\Sigma_n \equiv \mathbb{E}\left[S_n S_n^\top\right]$ *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 Σ_n is non-degenerate. Assumption B1(iv) mainly restricts the rate at which the dimension of ε_t grows to infinity. Under this assumption, we can verify the conditions in Theorem 2 and obtain a strong approximation for S_n, as stated by the following proposition.

Proposition B1. *Under Assumption B1, we have*

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

where $\widetilde{S}_n$ *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 B1 improves significantly the rate derived in Dehling (1983).

Proof of Proposition B1. *Step 1.* In this step, we verify the conditions of Theorem 2. Condition (iii) of Theorem 2 coincides with Assumption B1(iii). It remains to verify conditions (i) and (ii) of that theorem.

We first show that Assumption 2 holds for the $X_{n,t}$ array (i.e., condition (i) of Theorem 2). 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)),

$$\left\|\mathbb{E}\left[X_{n,t}^{(j)}|\mathcal{F}_{n,t-s}\right]\right\|_q \leq 6c_{\kappa,n} k_n^{-1/2} \varphi_s^{1/q-1/\kappa} \tag{B.1}$$

in the strong mixing case, and

$$\left\|\mathbb{E}\left[X_{n,t}^{(j)}|\mathcal{F}_{n,t-s}\right]\right\|_q \leq 2c_{\kappa,n} k_n^{-1/2} \varphi_s^{1-1/\kappa} \tag{B.2}$$

in the uniform mixing case. Therefore, $(X_{n,t})_{t=-\infty}^{\infty}$ is an L_q-mixingale array with $\bar{c}_n = 6c_{\kappa,n}$ and $\psi_s = \varphi_s^{1/q-1/\kappa}$ for the strong mixing case, and $\bar{c}_n = 2c_{\kappa,n}$ and $\psi_s = \varphi_s^{1-1/\kappa}$ 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 2, that is, the approximating martingale difference $X_{n,t}^*$ satisfies Assumption 1. Note that $X_{n,t}^* = k_n^{-1/2} \varepsilon_t^*$ where

$$\varepsilon_t^* \equiv \sum_{s=-\infty}^{\infty} \left\{\mathbb{E}\left[\varepsilon_{t+s}|\mathcal{F}_{n,t}\right] - \mathbb{E}\left[\varepsilon_{t+s}|\mathcal{F}_{n,t-1}\right]\right\}. \tag{B.3}$$

We denote the conditional covariance matrix of $X_{n,t}^*$ by

$$V_{n,t}^* = \mathbb{E}\left[X_{n,t}^* X_{n,t}^{*\top}|\mathcal{F}_{n,t-1}\right] = k_n^{-1} v_t^*,$$

where $v_t^* \equiv \mathbb{E}\left[\varepsilon_t^* \varepsilon_t^{*\top}|\mathcal{F}_{n,t-1}\right]$. Since ε_t is stationary, $(\varepsilon_t^*)_{t\geq 1}$ is also stationary. In particular,

$$k_n \mathbb{E}\left[V_{n,t}^*\right] = \mathbb{E}\left[v_t^*\right] = \Sigma_n^*. \tag{B.4}$$

Like (A.34), we can show that

$$\|\Sigma_n - \Sigma_n^*\|_S = O_p(c_{\kappa,n} m_n^{1/2} k_n^{-1/2} + c_{\kappa,n}^2 m_n k_n^{-1}) = o(1), \tag{B.5}$$

where the second equality is due to Assumption B1(iv). Hence, Assumption B1(iii) implies that the eigenvalues of Σ_n^* is bounded away from zero and from above. In view of (B.4), 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 1, with

$$r_n = c_{\kappa,n}^2 m_n k_n^{-1/2}. \tag{B.6}$$

Step 2. By the derivations in step 1, we can apply Theorem 2 to show that

$$\begin{aligned}\left\|S_n - \widetilde{S}_n\right\| &= O_p(c_{\kappa,n} m_n^{1/2} k_n^{-1/2}) + O_p(m_n^{1/2} r_n^{1/2} + (B_n^* m_n)^{1/3}) \\ &\quad + O_p(c_{\kappa,n} m_n k_n^{-1/2} + c_{\kappa,n}^2 m_n^{3/2} k_n^{-1}).\end{aligned}$$

Following the same argument as in Corollary 1, we deduce $B_n^* = O(c_{\kappa,n}^3 m_n^{3/2} k_n^{-1/2})$. Using this estimate and (B.6), we can simplify the error bound above as

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

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

S.B.2 Primitive conditions for Assumption 6

In this subsection, we illustrate how to verify Assumption 6 under the following primitive condition.

Assumption B2. *(i)* $(X_t^\top, u_t)_t$ *is a strictly stationary strong mixing process with mixing coefficient* $(\varphi_s)_{s=0}^\infty$ *satisfying* $\sum_{s=1}^\infty s^{r_2} \varphi_s^{(\kappa-4)/(5\kappa)}$ *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)* $\mathbb{E}\left[|u_t|^\kappa |X_t\right] \le C < \infty$ *almost surely for any t; (iv)* $\max_{1\le k\le m_n} \sup_{x\in\mathcal{X}} |p_k(x)| \le \zeta_n$ *where* ζ_n *is a non-decreasing positive sequence and* $\log(\zeta_n^L) = O(\log(m_n))$*; (v) there exist* $\rho_h > 0$ *and* $b_n^* \in \mathbb{R}^{m_n}$ *such that*

$$\sup_{x\in\mathcal{X}} \left|P(x)^\top b_n^* - h(x)\right| = O(m_n^{-\rho_h});$$

(vi) $\inf_{x\in\mathcal{X}} \|P(x)\| \ge c$ *for all* m_n *and some constant* $c > 0$*; (vii)* $(\zeta_n + m_n^{1/2}) M_n \zeta_n m_n n^{-1/2} + n^{1/2} m_n^{-\rho_h} + \zeta_n^6 m_n^5 n^{-1} = o(1)$ *and* $\zeta_n^2 m_n M_n^{-r_1\wedge r_2} = o(1)$ *where* r_1 *is the constant defined in Assumption 3.*

Assumption B2(i) imposes restrictions on the serial dependence of the data. Assumption B2(ii) is the same as Assumption 6(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 $P(x)^\top b_n^*$ with approximation error $O(m_n^{-\rho_h})$ 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.

Lemma B1. *Under Assumption B2, Assumption 6(iii) holds with* $\delta_{2,n} = \zeta_n m_n^{5/6} n^{-1/6}$.

Proof of Lemma B1. We use Proposition B1 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(iii) and B2(iv),

$$\sup_{t,j} \left\| \varepsilon_t^{(j)} \right\|_\kappa = \sup_{t,j} \| u_t p_j(X_t) \|_\kappa \leq \zeta_n \sup_t \| u_t \|_\kappa \leq C^{1/\kappa} \zeta_n,$$

which verifies Assumption B1(i) with $c_{\kappa,n} = C^{1/\kappa}\zeta_n$. Assumptions B1(ii), B1(iii) and B1(iv) are implied by Assumptions B2(i), B2(ii) and B2(vii), respectively. The assertion of Lemma B1 then follows from Proposition B1. *Q.E.D.*

Lemma B2. *Under Assumption B2, Assumption 6(iv) holds with* $\delta_{3,n} = m_n \zeta_n^2 n^{-1/2}$.

Proof of Lemma B2. 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,

$$\begin{aligned}
\mathbb{E}\left[\left\| \widehat{Q}_n - Q_n \right\|^2 \right] &= \mathbb{E}\left[\left\| n^{-1} \sum_{t=1}^{n} \left(P(X_t) P(X_t)^\top - \mathbb{E}\left[P(X_t) P(X_t)^\top \right] \right) \right\|^2 \right] \\
&= \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \mathbb{E}\left[\left(n^{-1} \sum_{t=1}^{n} \eta_t(j,k) \right)^2 \right] \\
&= n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=1}^{n} \mathbb{E}\left[\eta_t^2(j,k) \right] \\
&\quad + 2n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \mathbb{E}\left[\eta_t(j,k) \eta_s(j,k) \right].
\end{aligned} \tag{B.7}$$

By Assumptions B2(iv),

$$n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=1}^{n} \mathbb{E}\left[\eta_t^2(j,k) \right] \leq n^{-2} \sum_{j=1}^{m_n} \sum_{k=1}^{m_n} \sum_{t=1}^{n} \mathbb{E}\left[p_j(X_t)^2\, p_k(X_t)^2 \right] \leq m_n^2 \zeta_n^4 n^{-1}. \tag{B.8}$$

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 $(\varphi_s)_{s=0}^\infty$ 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),

$$\left| \mathbb{E}\left[\eta_t(j,k) \eta_s(j,k) \right] \right| \leq K \varphi_{t-s}^{1-2/\kappa} \left\| p_j(X_t) p_k(X_t) \right\|_\kappa \left\| p_j(X_s) p_k(X_s) \right\|_\kappa \leq K \varphi_{t-s}^{1-2/\kappa} \zeta_n^4. \tag{B.9}$$

By (B.9) and the summability condition of the mixing coefficients in Assumption B2(i),

$$
\begin{aligned}
& n^{-2}\sum_{j=1}^{m_n}\sum_{k=1}^{m_n}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left|\mathbb{E}\left[\eta_t(j,k)\eta_s(j,k)\right]\right| \\
& \quad \leq K m_n^2\zeta_n^4 n^{-2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\varphi_{t-s}^{1-2/\kappa} = O(m_n^2\zeta_n^4 n^{-1}).
\end{aligned} \tag{B.10}
$$

From (B.7), (B.8) and (B.10), we deduce $\mathbb{E}[\|\widehat{Q}_n - Q_n\|^2] = O(m_n^2\zeta_n^4 n^{-1})$, which readily implies the assertion of Lemma B2. *Q.E.D.*

Lemma B3. *Under Assumptions 3 and B2, Assumption 6(v) holds with*

$$
\delta_{4,n} = \zeta_n^2 m_n(M_n n^{-1/2} + M_n^{-r_1\wedge r_2}) + M_n\zeta_n m_n^{1-\rho_h} + M_n\zeta_n m_n^{3/2}n^{-1/2}.
$$

Proof of Lemma B3. *Step 1.* We use Theorem 4 to prove this lemma. In order to cast the setting into that of Theorem 4, we set $k_n = n$, $Z_t = (Y_t, X_t^\top)$, $\theta_0 = h(\cdot)$, $\hat{\theta}_n = \hat{h}_n(\cdot)$ and

$$
X_{n,t} = n^{-1/2}P(X_t)u_t, \quad \widehat{X}_{n,t} = n^{-1/2}P(X_t)\left(Y_t - \hat{h}_n(X_t)\right). \tag{B.11}
$$

In this step, we verify that the $X_{n,t}$ array satisfies Assumption 4. Under Assumption B2, we can use the same arguments in the proof of Proposition B1 to show that the array $(X_{n,t})$ satisfies Assumption 2 with $\bar{c}_n = 6C^{1/\kappa}\zeta_n$, $q = 5\kappa/(\kappa+1)$ and $\psi_s = \varphi_s^{(\kappa-4)/(5\kappa)}$. It remains to verify conditions (i)–(iv) in Assumption 4.

By (3.14), $\mathbb{E}[X_{n,t}] = 0$ for any t and any n. Moreover, by Assumption B2(i), $\mathbb{E}[X_{n,t}X_{n,t+j}] = n^{-1}\mathbb{E}\left[u_t u_{t+j}P(X_t)P(X_{t+j})^\top\right]$ only depends on n and j. Therefore, Assumption 4(i) holds. Let $\mathcal{F}_{n,t}$ be the σ-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

$$
\left\|\mathbb{E}\left[X_{n,t}^{(l)}X_{n,t+j}^{(k)}\middle|\mathcal{F}_{n,t}\right] - \mathbb{E}\left[X_{n,t}^{(l)}X_{n,t+j}^{(k)}\right]\right\|_2 \leq 6\varphi_s^{1/2-2/\kappa}\left\|X_{n,t}^{(l)}X_{n,t+j}^{(k)}\right\|_{\kappa/2}. \tag{B.12}
$$

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

$$
\left\|X_{n,t}^{(l)}X_{n,t+j}^{(k)}\right\|_{\kappa/2} \leq \|u_t u_{t+j}\|_{\kappa/2}\zeta_n^2 n^{-1} \leq C^{2/\kappa}\zeta_n^2 n^{-1} \leq \bar{c}_n^2 n^{-1}, \tag{B.13}
$$

which verifies Assumption 4(iii). Furthermore, this estimate and (B.12) imply that

$$
\left\|\mathbb{E}\left[X_{n,t}^{(l)}X_{n,t+j}^{(k)}\middle|\mathcal{F}_{n,t}\right] - \mathbb{E}\left[X_{n,t}^{(l)}X_{n,t+j}^{(k)}\right]\right\|_2 \leq 6C^{2/\kappa}\zeta_n^2\varphi_s^{1/2-2/\kappa}n^{-1} \leq \bar{c}_n^2 n^{-1}\varphi_s^{(\kappa-2)/(2\kappa)}. \tag{B.14}
$$

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

Finally, we verify Assumption 4(iv). Under Assumption B2(i), ψ_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 ψ_s that is not summable. Therefore,

$$
\sup_{s\geq 0} s\psi_s^2 \leq 1 + \max_{0\leq s\leq s_0} s\psi_s^2 < \infty.
$$

Further note that $\sum_{s=0}^{\infty} s^{r_2}\psi_s < \infty$ holds by Assumption B2(i). This verifies Assumption 4(iv).

Step 2. In this step, we finish the proof of Lemma B3 by verifying Assumption 5 for which we note from (B.11) 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),

$$
\begin{aligned}
n^{-1}\sum_{t=1}^{n}\left\|g(Z_n,\hat{h}_n)-g(Z_t,h)\right\|^2 &= n^{-1}\sum_{t=1}^{n}\left(\hat{h}_n(X_t)-h(X_t)\right)^2 P(X_t)^\top P(X_t) \\
&\le \zeta_n^2 m_n n^{-1}\sum_{t=1}^{n}\left(\hat{h}_n(X_t)-h_0(X_t)\right)^2. \qquad \text{(B.15)}
\end{aligned}
$$

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

$$
\lambda_{\min}^{-1}(\widehat{Q}_n) + \lambda_{\max}(\widehat{Q}_n) \le K, \quad \text{with probability approaching one.} \qquad \text{(B.16)}
$$

Define P_n, U_n, H_n and H_n^* as in the proof of Theorem 5. Like (A.68) and (A.70), we can show that

$$
\hat{b}_n - b_n^* = (P_n^\top P_n)^{-1}\left(P_n^\top U_n\right) + (P_n^\top P_n)^{-1}P_n^\top(H_n - H_n^*), \qquad \text{(B.17)}
$$

and $n^{-1/2}P_n^\top U_n = O_p(m_n^{1/2})$. Then, by (B.16),

$$
\left\|(P_n^\top P_n)^{-1}\left(P_n^\top U_n\right)\right\| \le \lambda_{\min}^{-1}(\widehat{Q}_n)\left\|n^{-1}P_n^\top U_n\right\| = O_p(m_n^{1/2}n^{-1/2}). \qquad \text{(B.18)}
$$

By Assumption B2(v) and (B.16),

$$
\left\|(P_n^\top P_n)^{-1}P_n^\top(H_n - H_n^*)\right\|^2 \le \lambda_{\min}^{-1}(\widehat{Q}_n)n^{-1}\left\|H_n - H_n^*\right\|^2 = O_p(m_n^{-2\rho_h}). \qquad \text{(B.19)}
$$

By (B.17), (B.18) and (B.19),

$$
\left\|\hat{b}_n - b_n^*\right\| = O_p(m_n^{1/2}n^{-1/2} + m_n^{-\rho_h}). \qquad \text{(B.20)}
$$

By Assumption B2(v), (B.16) and (B.20),

$$
\begin{aligned}
& n^{-1}\sum_{t=1}^{n}\left(\hat{h}_n(X_t)-h(X_t)\right)^2 \\
&\le 2n^{-1}\sum_{t=1}^{n}\left(\hat{h}_n(X_t)-P(X_t)^\top b_n^*\right)^2 + 2n^{-1}\sum_{t=1}^{n}\left(P(X_t)^\top b_n^* - h(X_t)\right)^2 \\
&\le 2\lambda_{\max}(\widehat{Q}_n)\left\|\hat{b}_n - b_n^*\right\|^2 + 2\sup_{x\in\mathcal{X}}\left|P(x)^\top b_n^* - h(x)\right|^2 \\
&= O_p(m_n^{-2\rho_h} + m_n n^{-1}).
\end{aligned} \qquad \text{(B.21)}
$$

Combined with (B.15), this estimate further implies that

$$
n^{-1}\sum_{t=1}^{n}\left\|g(Z_t,\hat{h}_n)-g(Z_t,h)\right\|^2 = O_p(\zeta_n^2 m_n^{1-2\rho_h} + \zeta_n^2 m_n^2 n^{-1}), \qquad \text{(B.22)}
$$

which verifies Assumption 5(i) with $\delta_{\theta,n} = \zeta_n m_n^{1/2-\rho_h} + \zeta_n m_n n^{-1/2}$. By Assumption B2(ii,iii),

$$\|g(Z_t,h)\|_2^2 = \mathbb{E}\left[u_t^2 P(X_t)^\top P(X_t)\right] \leq C^{2/\kappa}\mathrm{Tr}\left(Q_n\right) \leq K m_n,$$

which implies Assumption 5(ii). This finishes the proof. *Q.E.D.*

Proposition B2. *Assumption 6 holds under Assumptions 3 and B2.*

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

$$\sup_{x\in\mathcal{X}} \frac{n^{1/2}\left|h(x) - P(x)^\top b_n^*\right|}{\|P(x)\|} \leq O(n^{1/2} m_n^{-\rho_h}). \tag{B.23}$$

Therefore, Assumption 6(i) holds with $\delta_{1,n} = n^{1/2} m_n^{-\rho_h}$, where $\delta_{1,n} = o(1)$ under Assumption B2(vii). Assumption 6(ii) is directly assumed in Assumption B2(ii). Assumptions 6(iii), 6(iv) and 6(v) have been verified in Lemma B1, Lemma B2 and Lemma B3, respectively; $\delta_{j,n} = o(1)$, $j \in \{2,3,4\}$, holds because of Assumption B2(vii). *Q.E.D.*

S.B.3 Time-series inference for intersection bounds

As mentioned in footnote 16 of the main text, we can verify the high-level conditions in Lemma 5 of Chernozhukov, Lee, and Rosen (2013), so as to extend their series-based inference for intersection bounds to the time-series setting. Proposition B3, below, provides the details. Note that we only need to verify Conditions NS(i)(a) and NS(ii) in Chernozhukov, Lee, and Rosen (2013), because the other conditions do not involve further complications resulted from the time-series extension.

Proposition B3. *Suppose Assumption 6 holds. If we futher have*

$$\delta_{1,n} + \delta_{2,n} + m_n^{1/2}\delta_{3,n} = o(1/\log(n)) \tag{B.24}$$

and

$$m_n^{1/2}(\delta_{3,n} + \delta_{4,n}) = n^{-b} \tag{B.25}$$

where $b > 0$ is a constant, then Conditions NS(i)(a) and NS(ii) in Chernozhukov, Lee, and Rosen (2013) hold.

Proof of Proposition B3. By (A.65), (A.81) and Assumption 6(ii,iii),

$$\frac{n^{1/2}P(x)^\top(\hat{b}_n - b_n^*)}{\sigma_n(x)} = \frac{P(x)^\top Q_n^{-1}\widetilde{N}_n}{\sigma_n(x)} + O_{pu}(\delta_{1,n} + \delta_{2,n} + m_n^{1/2}\delta_{3,n}), \tag{B.26}$$

where $\widetilde{N}_n \sim \mathcal{N}(0, A_n)$ and $O_{pu}(\cdot)$ denotes a uniformly (in x) stochastically bounded sequence. By (A.83) and (B.26),

$$\begin{aligned}\frac{n^{1/2}(\hat{h}_n(x) - h(x))}{\sigma_n(x)} &= \frac{P(x)^\top Q_n^{-1}\widetilde{N}_n}{\sigma_n(x)} + O_{pu}(\delta_{1,n} + \delta_{2,n} + m_n^{1/2}\delta_{3,n}) \\ &= \frac{P(x)^\top \Sigma_n^{1/2}\widetilde{N}_n^*}{\sigma_n(x)} + O_{pu}(\delta_{1,n} + \delta_{2,n} + m_n^{1/2}\delta_{3,n}),\end{aligned}$$

where $\widetilde{N}_n^* \equiv A_n^{-1/2}\widetilde{N}_n$ is an m_n-dimensional standard normal random vector. Condition NS(i)(a) in Chernozhukov, Lee, and Rosen (2013) then follows from (B.24).

By (A.74) and the relation between the spectral norm and the Frobenius norm of matrices,

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

Condition NS(ii) in Chernozhukov, Lee, and Rosen (2013) then follows from (B.25). *Q.E.D.*

S.B.4 Technical derivations of the conditional moment restriction in the search and matching model

In this appendix, we derive the conditional moment restriction (4.9) in the main text. Recall that the equilibrium is characterized by the following Bellman equations:

$$J_p = p - w_p + \delta(1-s)\mathbb{E}_p\left[J_{p'}\right], \tag{B.27}$$

$$V_p = -c_p + \delta q(\theta_p)\mathbb{E}_p\left[J_{p'}\right], \tag{B.28}$$

$$U_p = z + \delta\left\{f(\theta_p)\mathbb{E}_p\left[W_{p'}\right] + (1 - f(\theta_p))\mathbb{E}_p\left[U_{p'}\right]\right\}, \tag{B.29}$$

$$W_p = w_p + \delta\left\{(1-s)\mathbb{E}_p\left[W_{p'}\right] + s\mathbb{E}_p\left[U_{p'}\right]\right\}, \tag{B.30}$$

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

$$J_p = (W_p - U_p)(1-\beta)/\beta. \tag{B.31}$$

Taking a difference between (B.29) and (B.30) yields

$$W_p - U_p = w_p - z + \delta(1 - s - f(\theta_p))\mathbb{E}_p\left[W_{p'} - U_{p'}\right]. \tag{B.32}$$

Combining (B.32) with (B.31), we derive

$$J_p = \frac{1-\beta}{\beta}(w_p - z) + \delta(1 - s - f(\theta_p))\mathbb{E}_p\left[J_{p'}\right]. \tag{B.33}$$

From (B.27) and (B.33), we can solve for the wage function

$$w_p = \beta p + (1-\beta)z + \beta\delta f(\theta_p)\mathbb{E}_p\left[J_{p'}\right]. \tag{B.34}$$

Note that the free entry condition implies

$$\delta q\left(\theta_p\right)\mathbb{E}_p\left[J_{p'}\right]-c_p=0. \tag{B.35}$$

Since $f\left(\theta\right)/q\left(\theta\right)=\theta$, we can rewrite (B.34) as

$$w_p=\beta p+\left(1-\beta\right)z+\beta\theta_p c_p. \tag{B.36}$$

We can rewrite (B.35) as $\delta\mathbb{E}_p\left[J_{p'}\right]=c_p/q\left(\theta_p\right)$. Plugging this and (B.36) into (B.27), we deduce

$$J_p=\left(1-\beta\right)\left(p-z\right)-\beta\theta_p c_p+\left(1-s\right)\frac{c_p}{q\left(\theta_p\right)}. \tag{B.37}$$

Finally, plugging (B.37) into (B.35) yields

$$\delta q\left(\theta_p\right)\mathbb{E}_p\left[\left(1-\beta\right)\left(p'-z\right)-\beta\theta_{p'}c_{p'}+\left(1-s\right)\frac{c_{p'}}{q\left(\theta_{p'}\right)}\right]-c_p=0,$$

as claimed in (4.8) of the main text. The conditional moment restriction (4.9) is then obtained by replacing θ and p with observed data.

References

Andrews, D. W. K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59(2), 307–345.

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015): "Some new asymptotic theory for least squares series: Pointwise and uniform results," *Journal of Econometrics*, 186(2), 345 – 366, High Dimensional Problems in Econometrics.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, chap. 76. Elsevier, 1 edn.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press.

Dehling, H. (1983): "Limit theorems for sums of weakly dependent Banach space valued random variables," *Wahrscheinlichkeitstheorie verw, Gebiete 63,*.

Newey, W. K. (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79(1), 147 – 168.

Pollard, D. (2001): *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.