Journal of Econometrics 219 (2020) 38–51

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Uniform nonparametric inference for time series[☆]

Jia Li[a,*], Zhipeng Liao[b]

[a] Department of Economics, Duke University, Durham, NC 27708, United States of America
[b] Department of Economics, UCLA, Log Angeles, CA 90095, United States of America

ARTICLE INFO

Article history:
Received 14 March 2019
Received in revised form 14 May 2019
Accepted 21 September 2019
Available online 28 April 2020

JEL classification:
C14
C22

Keywords:
Martingale difference
Mixingale
Series estimation
Specification test
Strong approximation
Uniform inference

ABSTRACT

This paper provides the first result for the uniform inference based on nonparametric series estimators in a general time-series setting. We develop a strong approximation theory for sample averages of mixingales with dimensions growing with the sample size. We use this result to justify the asymptotic validity of a uniform confidence band for series estimators and show that it can also be used to conduct nonparametric specification test for conditional moment restrictions. New results on the validity of heteroskedasticity and autocorrelation consistent (HAC) estimators with increasing dimension are established for making feasible inference. An empirical application on the unemployment volatility puzzle for the search and matching model is provided as an illustration.

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

☆ This paper is a substantially revised version of our working paper previously circulated as "Uniform Nonparametric Series Inference for Dependent Data with an Application to the Search and Matching Model". Suggestions from a Co-Editor and anonymous referees have greatly improved the paper. We thank Don Andrews, Tim Bollerslev, Xiaohong Chen, Denis Chetverikov, Cosmin Ilut, Kyle Jurado, Roger Koenker, Andrea Lanteri, Taisuke Otsu, Andrew Patton, Peter C.B. Phillips, Adam Rosen, George Tauchen and seminar and conference participants at Cornell, Duke, London School of Economics, Singapore Management University, UCL, UCLA, the 2018 CEME conference, the 2018 California econometrics conference and the Cowles Foundation conference on celebration of Peter C.B. Phillips' forty years at Yale for their comments. Liao's research was partially supported by National Science Foundation, United States of America Grant SES-1628889.

* Corresponding author.
E-mail addresses: jl410@duke.edu (J. Li), zhipeng.liao@econ.ecla.edu (Z. Liao).

https://doi.org/10.1016/j.jeconom.2019.09.011
0304-4076/

the inference of "irregular" functionals of the nonparametric component, with the leading example being the *pointwise* inference for the unknown function, where the irregular functional evaluates the function at a given point. The resulting estimator has a slower nonparametric rate of convergence.

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

Recently, Chernozhukov et al. (2013c) and Belloni et al. (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 data vector with increasing dimension. The uniform series inference theory of Chernozhukov et al. (2013c) and Belloni et al. (2015) relies on this type of coupling, and hence, are restricted to cross-sectional applications with independent data.[1]

Set against this background, our initial contribution (see Section 2) is to develop a uniform inference theory for series estimators in time-series applications. To do so, we establish a novel strong approximation (i.e., coupling) result for general heterogeneous martingale difference arrays, which arise routinely from empirical applications on dynamic stochastic equilibrium models. Compared with the classical Yurinskii coupling for independent data, the key complication in the martingale-difference setting stems from the stochastic volatility of time series data, and we address this complication using a novel martingale technique. Armed with this coupling result, we then establish a uniform inference theory for series estimators in the time-series setting. We also show that the uniform inference can be conveniently used for nonparametrically testing conditional moment equalities implied by Euler or Bellman equations in dynamic stochastic equilibrium models.

In Section 3, we extend the aforementioned baseline theory to allow for more general forms of dependence. These additional results are new to the literature, which further set our analysis apart from existing work. First, we extend our martingale-difference coupling result to general mixingales by developing a martingale approximation technique with increasing dimension. It is well known that mixingales form a far more general class of processes than those characterized by various mixing concepts. For example, Andrews (1984, 1985) showed constructively that even nearly independent triangular arrays are not strongly mixing; as a result of the ranking of mixing coefficients, they are not ρ-mixing or β-mixing, either.[2] By contrast, mixingales include martingale differences, ARMA processes, linear processes, various mixing and near-epoch dependent series as special cases. Therefore, our contribution is not limited to merely allowing for "some" specific form of dependence, but rather to provide a very general theory for essentially all types of dependence used in time-series econometrics. The general limit theorems developed here should be broadly useful for future research in nonparametric time-series settings. Second, in order to conduct feasible inference in this general setting, we prove the validity of classical HAC estimators for long-run covariance matrices but in the current nonstandard setting with growing dimension. This result is of independent econometric interest and may be useful in other inference problems as well.

As a concrete empirical illustration of the proposed method, we study the unemployment volatility puzzle within the standard search and matching model (Pissarides, 1985; Mortensen and Pissarides, 1994; Pissarides, 2000). In an influential paper, Shimer (2005) shows that the standard Mortensen–Pissarides model, when calibrated in the conventional way, generates unemployment volatility that is far lower than the empirical estimate. Various modifications to the standard model have been proposed to address this puzzle; see Shimer (2004), Hall (2005), Mortensen and Nagypál (2017), Hall and Milgrom (2008), Pissarides (2009), and references therein. Hagedorn and Manovskii (2008), on the other hand, take a different route and show that the standard model actually can generate high levels of unemployment volatility using their alternative calibration strategy. The plausibility of their alternative calibration remains a contentious issue in the literature (see Hornstein et al., 2005). To shed some light on this debate from an econometric perspective, we derive a conditional moment restriction from the equilibrium Bellman equations. We then test whether this restriction holds or not at the parameter values calibrated by Hagedorn and Manovskii (2008) using the proposed uniform inference method. The nonparametric specification test strongly rejects the hypothesis that these calibrated values are compatible with the equilibrium conditional moment restriction, and hence suggests that modifications to the standard Mortensen–Pissarides model are necessary for a better understanding of the cyclicality of unemployment. Constructively, we compute

1 Yurinskii's coupling concerns the strong approximation of a high-dimensional vector under the Euclidean distance. Chernozhukov et al. (2014) establish a strong approximation for the largest entry of a high-dimensional vector under a more general setting.

2 On the other hand, Andrews's examples are linear processes, which are special cases of mixingales. In probability theory, counterexamples of this kind can be traced back to Ibragimov and Linnik (1971) and Chernick (1981). See also Davidson (1994) for a comprehensive review on various dependence concepts used in econometrics.

the Anderson–Rubin confidence set of parameter values which the test does not reject, which is informative about the "admissible" range of parameters for future research.

The present paper is related to several strands of literature in econometrics and statistics. 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 and Shen (1998), He and Shao (2000), Chen (2007), Chen and Pouzo (2012), Chen and Liao (2014), Chen et al. (2014), Chen and Pouzo (2015), Chen and Christensen (2015) and Hansen (2015) and many references therein. In particular, Chen and Christensen (2015) establish (under weak conditions) the minimax sup-norm rate for time-series series least-square regression and linear and nonlinear functionals inference, but they do not study uniform inference. Chernozhukov et al. (2013c, 2014) and Belloni et al. (2015) studied uniform series inference for independent data.[3] Unlike all aforementioned work, our econometric focus is on the *uniform* series inference for *general time-series* data, which extends this large and burgeoning literature in a new dimension.

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 dimension. This result is of independent interest and should be useful for other types of time-series inference as well.

Finally, on the statistical side, our strong approximation results for heterogeneous martingale difference arrays and (more generally) mixingales are related to the literature on coupling with increasing dimension; in particular, we extend Yurinskii's coupling (Yurinskii, 1978) from the independent data setting to a general time-series setting. The recent work of Chernozhukov et al. (2013c) relies on Yurinskii's coupling and our (much) more general coupling theory may be used to extend their intersection-bound-based inference to a general time series setting. Chernozhukov et al. (2013a) constructed a strong approximation for the largest entry of a vector with increasing dimension and Belloni et al. (2015) applied this coupling to conduct uniform series inference. This alternative form of coupling is implied by Yurinskii's coupling but can be obtained under weaker restrictions on the growth rate of the dimensionality. We focus purposefully on Yurinskii-type coupling so as to cover a wide range of empirical applications, leaving the technical pursuit of weaker growth conditions on the dimensionality to future research. There has been limited research on coupling with increasing dimension in the time-series setting. Chernozhukov et al. (2013b) establish the strong approximation for the largest entry of a β-mixing sequence.[4] As mentioned before, our coupling result is valid for general heterogeneous mixingales, which are far more general than mixing processes, and is thus free of the well-known critique of Andrews (1984, 1985).[5] 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).[6]

The paper is organized as follows. Section 2 represents the baseline econometric theory, which is further extended in Section 3 . The empirical application is given in Section 4. Section 5 concludes. All proofs for our theoretical results are in the supplemental appendix of this paper.

Notations. For any real matrix A, we use $\|A\|$ and $\|A\|_S$ to denote its Frobenius norm and spectral norm, respectively. We use $a^{(j)}$ to denote the 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. Baseline results on uniform inference

In order to streamline the discussion, we start in this section with our baseline theoretical results for uniform nonparametric time-series inference. Further extensions, which are substantially more general, are gathered in Section 3. Section 2.1 presents the strong approximation theorem for heterogeneous martingale differences. The uniform inference theory for series estimators in the time-series setting is presented in Section 2.2. Section 2.3 provides further results on how to use this uniform inference theory for testing conditional moment restrictions.

3 In a recent paper, Chen and Christensen (2018) establish the minimax sup-norm rate and strong approximation in nonparametric instrumental variables (NPIV) problems in the i.i.d. setting. We do not consider time-series NPIV problems in this paper, which may be interesting for future research.

4 Zhang and Wu (2017) establish a similar coupling based on a notion of dependence obtained from stationary nonlinear systems (Wu, 2005). Also see Zhang and Cheng (2018) for a similar result under an alternative notion of physical dependence.

5 Technically speaking, the martingale-based technique developed here is very different from the "large-block–small-block" technique employed in Chernozhukov et al. (2013b), and it is necessitated by the distinct dependence structure studied in the present paper.

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

2.1. 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; a further extension to mixingales is in Section 3.1. We single out the result for martingale differences mainly because they arise routinely from dynamic stochastic equilibrium models that are equipped with information filtrations. Hence, this result is directly applicable in many economic applications. In addition, the inference for martingale differences does not involve the nonstandard HAC estimation with increasing dimension, and hence, permits a relatively simple implementation.

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We consider an m_n-dimensional square-integrable martingale difference array $(X_{n,t})_{1\leq t\leq k_n, n\geq 1}$ with respect to a filtration $(\mathcal{F}_{n,t})_{1\leq t\leq k_n, n\geq 1}$, where $k_n \to \infty$ as $n \to \infty$. That is, $X_{n,t}$ is $\mathcal{F}_{n,t}$ -measurable with finite second moment and $\mathbb{E}\left[X_{n,t}|\mathcal{F}_{n,t-1}\right] = 0$. 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

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

It is instructive to note that for typical applications k_n is often the sample size n and the array $X_{n,t}$ represents a $n^{-1/2}$-normalized version of some time series, so the order of magnitude of $V_{n,t}$ is n^{-1}. For simplicity, we denote $\Sigma_n \equiv \Sigma_{n,k_n}$ in the sequel.

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

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

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

$$S_n \xrightarrow{d} \mathcal{N}\left(0, \Sigma\right), \tag{2.1}$$

where $\Sigma = \lim_{n\to\infty} \Sigma_n$. In the present paper, however, we are mainly interested in the case with $m_n \to \infty$, which is particularly relevant for analyzing the nonparametric series estimation with growing number of approximating functions. We aim to construct a coupling sequence $\widetilde{S}_n \sim \mathcal{N}\left(0, \Sigma_n\right)$ such that $\|S_n - \widetilde{S}_n\| = O_p\left(\delta_n\right)$ for some $\delta_n \to 0$. The following assumption is needed.

Assumption 1. Suppose (i) the eigenvalues of $k_n\mathbb{E}\left[V_{n,t}\right]$ 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 $h_n/k_n \to 1$,

$$\left\| \sum_{t=1}^{h_n} V_{n,t} - \Sigma_{n,h_n} \right\|_S = O_p(r_n), \tag{2.2}$$

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) requires the conditional covariance of the martingale S_n (i.e., $\sum_{t=1}^{h_n} V_{n,t}$) to be close to its unconditional mean. This condition can be verified easily under mild weak-dependence type assumptions on the conditional covariance $V_{n,t}$.

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

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

$$\left\|S_n - \widetilde{S}_n\right\| = O_p(m_n^{1/2}r_n^{1/2} + (B_n m_n)^{1/3}), \tag{2.3}$$

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

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 $\left\|S_n - S_n^*\right\| = O_p(m_n^{1/2}r_n^{1/2})$. This approximation step is not needed in the conventional setting with independent data, because in the latter case the conditional covariance process $V_{n,t}$ is nonrandom. Hence, the $O_p(m_n^{1/2}r_n^{1/2})$ error term is the "cost" for accommodating stochastic volatility.[7] In the second step, we establish a strong approximation for S_n^*.

[7] 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 (2.2) 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.

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 Strassen's theorem for establishing the strong approximation. The resulting approximation error is $O_p((B_n m_n)^{1/3})$, which is essentially the rate of the classical Yurinskii's coupling for independent data. The typical magnitude of B_n is $O(k_n^{-1/2} m_n^{3/2})$ and, correspondingly, $(B_n m_n)^{1/3} = O(k_n^{-1/6} m_n^{5/6})$.

Finally, we note that an alternative form of strong approximation could be constructed, that is, a coupling for the largest entry of the vector S_n instead of the vector itself; this alternative type of coupling is weaker than Yurinskii's coupling (i.e., the former is implied by the latter), and may require less restrictions on the growth rate of the dimension m_n. That being said, we focus intentionally on the Yurinskii-type coupling in this paper for two related reasons. One is that establishing this type of coupling in the general time-series setting is evidently of its own independent theoretical importance and sets a relevant benchmark for future work. Indeed, some inference problems may actually require the strong approximation for the entire vector instead of just its largest entry. The other reason is more "practical:" our results allow us to extend existing inference method that are based on Yurinskii-type coupling to a general time-series setting with minimum additional cost. The pursuit of weaker conditions on the growth rate of m_n is secondary to our main focus in the current paper, and is left for future research.

2.2. Uniform inference for nonparametric series regressions

In this subsection, we apply the coupling theorem above to develop an asymptotic theory for conducting uniform nonparametric inference based on series estimation. Specifically, we consider the following nonparametric regression model: for $1 \leq t \leq n$,

$$Y_t = h(X_t) + u_t \tag{2.4}$$

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

$$\mathbb{E}[u_t|\mathcal{F}_{t-1}] = 0, \tag{2.5}$$

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.

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 approximating 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,

$$\hat{b}_n \equiv \left(\sum_{t=1}^{n} P(X_t) P(X_t)^\top\right)^{-1} \left(\sum_{t=1}^{n} P(X_t) Y_t\right). \tag{2.6}$$

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.

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 $n^{1/2}(\hat{h}_n(x) - h(x))$ is thus

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

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

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

Since u_t forms a martingale difference sequence, $A_n = n^{-1} \sum_{t=1}^{n} \mathbb{E}[u_t^2 P(X_t)P(X_t)^\top]$ and it can be estimated by

$$\widehat{A}_n \equiv n^{-1} \sum_{t=1}^{n} \widehat{u}_t^2 P(X_t)P(X_t)^\top, \quad \text{where} \quad \widehat{u}_t = Y_t - \hat{h}_n(X_t). \tag{2.7}$$

More generally, if we weaken the mean independence condition (2.5) to $\mathbb{E}[u_t|X_t] = 0$, then A_n is generally a (pre-asymptotic) long-run covariance matrix, and $\widehat{A}_n$ should be the corresponding HAC estimator. This extension is discussed in detail in Section 3.2.

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

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

Under some regularity conditions, we shall show (see Theorem 2) that the "sup-t" statistic

$$\widehat{T}_n \equiv \sup_{x\in\mathcal{X}} \left| \frac{n^{1/2}\left(\hat{h}_n(x) - h(x)\right)}{\hat{\sigma}_n(x)} \right|, \tag{2.8}$$

can be (strongly) approximated by

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

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 $h(\cdot)$. Algorithm 1, summarizes the implementation details.

Algorithm 1 (*Uniform confidence band construction*).
Step 1. Draw m_n-dimensional standard normal vectors ξ_n repeatedly and compute

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

Step 2. Set $cv_{n,\alpha}$ as the $1-\alpha$ quantile of $\widetilde{T}_n^*$ in the simulated sample.
Step 3. Report $\widehat{L}_n(x) = \hat{h}_n(x) - n^{-1/2} cv_{n,\alpha} \hat{\sigma}_n(x)$ and $\widehat{U}_n(x) = \hat{h}_n(x) + n^{-1/2} cv_{n,\alpha} \hat{\sigma}_n(x)$ as the $(1-\alpha)$-level uniform confidence band for $h(\cdot)$. □

We are now ready to present the asymptotic theory that justifies the validity of the confidence band described in the algorithm above. Although we consider the case with martingale-difference u_t error term here, the theory is valid in much more general settings as we show in Section 3. To facilitate our later extensions, we collect the key ingredients of the theorem in the following high-level assumption. These conditions are either standard in the series estimation literature or can be verified using the limit theorems developed in the current paper.[8] 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 2. For each $j = 1, \ldots, 4$, let $\delta_{j,n} = o(1)$ be a positive sequence. Suppose: (i) $\log(\zeta_n^L) = O(\log(m_n))$ and there exists a sequence $(b_n^*)_{n\geq 1}$ of m_n-dimensional constant vectors such that

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

(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

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

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

A few remarks on Assumption 2 are in order. Conditions (i) and (ii) are fairly standard in series estimation; see, for example, Andrews (1991a), Newey (1997), Chen (2007) and Belloni et al. (2015).[9] In particular, condition (i) specifies the precision for approximating the unknown function $h(\cdot)$ via approximating functions, for which comprehensive results are available from numerical approximation theory. When u_t is a martingale difference sequence, $X_{n,t} = n^{-1/2}P(X_t)u_t$ forms a martingale difference array, so the strong approximation in condition (iii) can be verified by using Theorem 1. More generally, this condition can be verified by using Theorem 4 for mixingales. Conditions (iv) and (v) pertain to the convergence rates of $\widehat{Q}_n$ and $\widehat{A}_n$. In Section 3.2, we will develop convergence-rate results for $\widehat{A}_n$ in a more general setting with HAC estimation. The conditions regarding $\widehat{Q}_n$ can be verified in a similar, actually simpler, way.[10]

8 In Supplemental Appendix S.B.4, we verify these high-level conditions under more primitive ones.

9 For some approximating functions such as power series, condition (ii) holds under certain nonsingular transformation on the vector approximating functions, i.e., $BP(\cdot)$, where B is some non-singular constant matrix. Since the nonparametric series estimator $\hat{h}_n(\cdot)$ is invariant to any nonsingular transformation of $P(\cdot)$, we do not distinguish between $BP(\cdot)$ and $P(\cdot)$ throughout this paper for notational convenience.

10 In the special case when u_t forms a martingale difference sequence, and the data series are strictly stationary and β-mixing, conditions (iv) and (v) can be directly verified by the theory developed in Chen and Christensen (2015).

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

Theorem 2. *The following statements hold under Assumption 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 $1-\alpha$:

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

2.3. Specification test for conditional moment restrictions

The uniform inference method developed in Section 2.2 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

$$\mathbb{E}\left[g(Y_t^*, \gamma_0)|X_t\right] = 0, \tag{2.9}$$

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

$$Y_t = g(Y_t^*, \gamma_0),\ h(x) = \mathbb{E}\left[Y_t|X_t = x\right] \text{ and } u_t = Y_t - \mathbb{E}\left[Y_t|X_t\right]. \tag{2.10}$$

The test for (2.9) can then be carried out by examining whether the uniform confidence bound $[\widehat{L}_n(\cdot), \widehat{U}_n(\cdot)]$ covers the zero function.[11] Anderson–Rubin type confidence sets for γ_0 can also be constructed by inverting the tests.

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 (2.9) 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.[12]

Equipped with $\hat{\gamma}_n$, we can implement the econometric procedure described in Section 2.2, except that we take Y_t as the "generated" variable $g\left(Y_t^*, \hat{\gamma}_n\right)$. More precisely, we set

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

$\hat{h}_n(x) = P(x)^\top \hat{b}_n$, $\widehat{u}_t = g\left(Y_t^*, \hat{\gamma}_n\right) - \hat{h}_n(X_t)$ and then define $\hat{\sigma}_n(x)$ similarly as in Section 2.2. In the theory presented below, we aim to provide sufficient conditions such that replacing γ_0 with $\hat{\gamma}_n$ leads to negligible errors. The intuition is straightforward: the proxy error in the finite-dimensional estimator $\hat{\gamma}_n$ shrinks at a fast parametric rate, and hence, is asymptotically dominated by the relatively large statistical noise in the nonparametric test.[13] We formalize this intuition with a few assumptions.

Assumption 3. Conditions (i)–(iv) of Assumption 2 hold with $h(x) = \mathbb{E}[g(Y_t^*, \gamma_0)|X_t = x]$ and $u_t = g(Y_t^*, \gamma_0) - h(X_t)$, condition (v) of Assumption 2 holds for $\widehat{A}_n$ 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 3 allows us to cast the testing problem into the nonparametric regression setting of Section 2.2. These conditions can be verified in the same way as discussed above. However, this assumption is not enough for our analysis because condition (iii) pertains only to the strong approximation of the infeasible estimator defined using $g\left(Y_t^*, \gamma_0\right)$ 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 $g_\gamma(\cdot)$ and $g_{\gamma\gamma}(\cdot)$ to denote the first and the second partial derivatives of $g(y, \gamma)$ with respect to γ, and we set

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

---

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

12 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.

13 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.

Assumption 4. 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 $\mathbb{E}[\left\| g_\gamma(Y_t^*,\gamma_0) \right\|^2]$ 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 4 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. The same techniques used in the verification of Assumption 2 can be applied to verify these conditions as well.

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

Algorithm 2. (Specification Test of Conditional Moment Restrictions)
Step 1. Implement Algorithm 1 with $Y_t = g(Y_t^*, \hat{\gamma}_n)$ and obtain the sup-t statistic $\widehat{T}_n$ and the critical value $cv_{n,\alpha}$.
Step 2. Reject the null hypothesis (2.9) at significance level α if $\widehat{T}_n > cv_{n,\alpha}$. $\square$

Theorem 3. *Suppose that Assumptions 3 and 4 hold. Then under the null hypothesis (2.9), 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.*

Comment. Theorem 3 shows that the proposed specification test controls size under the null hypothesis and is consistent against fixed alternatives. It does not, however, characterize the test's local power property, which may be an interesting theoretical question for future research.

3. Extensions

We extend our baseline results in the previous section in various directions, which further set our analysis apart from the prior literature. Section 3.1 extends the martingale-difference strong approximation result towards the much more general class of mixingales. Section 3.2 establishes the validity of HAC estimators for mixingale data in the setting with increasing dimension.

3.1. Strong approximation for mixingales

Theorem 1 is restrictive for some time-series applications because martingale differences are serially uncorrelated. In this subsection, we extend that baseline coupling result towards mixingale processes by using a martingale approximation with increasing dimension. Mixingales form a very general class of models, including martingale differences, linear processes and various types of mixing and near-epoch dependent processes as special cases, and naturally allow for data heterogeneity. In particular, the mixingale concept is substantially more general than various mixing concepts (see, e.g., Andrews, 1984, 1985 for counterexamples, and White and Gallant, 1988; Davidson, 1994; Pötscher and Prucha, 2013 for comprehensive reviews) and readily accommodate most (if not all) applications in time series econometrics.

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

$$\left\| \mathbb{E}[X_{n,t}^{(j)} | \mathcal{F}_{n,t-k}] \right\|_q \le c_{n,t}\psi_k, \quad \left\| X_{n,t}^{(j)} - \mathbb{E}[X_{n,t}^{(j)} | \mathcal{F}_{n,t+k}] \right\|_q \le c_{n,t}\psi_{k+1}, \tag{3.1}$$

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 $k_n^{1/2} X_{n,t}$.

Assumption 5. The array $(X_{n,t})$ satisfies (3.1) for some $q \ge 3$. Moreover, for some positive sequence $\bar{c}_n$, $\sup_t \left| c_{n,t} \right| \le \bar{c}_n k_n^{-1/2} = O(1)$ and $\sum_{k\ge 0} \psi_k < \infty$.

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

$$X_{n,t} = X_{n,t}^* + \widetilde{X}_{n,t} - \widetilde{X}_{n,t+1} \tag{3.2}$$

where $X_{n,t}^{*} \equiv \sum_{s=-\infty}^{\infty}\left\{\mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t}\right]-\mathbb{E}\left[X_{n,t+s}|\mathcal{F}_{n,t-1}\right]\right\}$ forms a martingale difference and the "residual" variable $\widetilde{X}_{n,t}$ satisfies $\sup_{j,t}\|\widetilde{X}_{n,t}^{(j)}\|_2 = O(\bar{c}_n k_n^{-1/2})$.[14] This representation further permits an approximation of S_n via the martingale $S_n^{*}=\sum_{t=1}^{k_n}X_{n,t}^{*}$, that is,

$$\left\|S_n-S_n^{*}\right\|_2=\left\|\widetilde{X}_{n,1}-\widetilde{X}_{n,k_n+1}\right\|_2=O(\bar{c}_n m_n^{1/2}k_n^{-1/2}). \tag{3.3}$$

In the typical case with $\bar{c}_n = O(1)$, the approximation error in (3.3) is negligible as soon as the dimension m_n grows at a slower rate than k_n. Consequently, a strong approximation for the martingale S_n^{*} (as described in Theorem 1) is also a strong approximation for S_n. Theorem 4, formalizes this result.

Theorem 4. *Suppose (i) Assumption 5 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*

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

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

Comment. 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 $O_p(\bar{c}_n m_n k_n^{-1/2}+\bar{c}_n^2 m_n^{3/2}k_n^{-1})$.

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

3.2. HAC estimation with increasing dimension

We have shown in Theorem 4 the strong approximation for the statistic $\sum_{t=1}^{n}X_{n,t}$ in a general time-series setting. In this subsection, we establish the asymptotic validity of a class of HAC estimators for its (long-run) covariance matrix $\Sigma_n = Var(\sum_{t=1}^{n}X_{n,t})$ that is needed for conducting feasible inference. This result can be used to verify Assumption 2(v) for constructing uniform confidence bands. Compared with the conventional setting on HAC estimation (see, e.g., Hannan, 1970; Newey and West, 1987; Andrews, 1991b; Hansen, 1992; de Jong and Davidson, 2000, etc.), the main difference in our analysis is to allow the dimension m_n to diverge asymptotically.[15] Since the HAC estimation theory in the current setting with increasing dimension is clearly of independent interest and may be used in other types of problems, we aim to build the theory in a general setting.

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

$$\widetilde{\Gamma}_{X,n}(s)\equiv\sum_{t=1}^{k_n-s}X_{n,t}X_{n,t+s}^{\top} \tag{3.5}$$

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

$$\widetilde{\Sigma}_n\equiv\sum_{s=-k_n+1}^{k_n-1}\mathcal{K}(s/M_n)\,\widetilde{\Gamma}_{X,n}(s) \tag{3.6}$$

where $\mathcal{K}(\cdot)$ is a kernel smoothing function and M_n is a bandwidth parameter that satisfies $M_n\to\infty$ as $n\to\infty$. The kernel function satisfies the following standard assumption.

Assumption 6. (i) $\mathcal{K}(\cdot)$ is bounded, Lebesgue-integrable, symmetric and continuous at zero with $\mathcal{K}(0)=1$; (ii) for some constants $C\in\mathbb{R}$ and $r_1\in(0,\infty]$, $\lim_{x\to0}(1-\mathcal{K}(x))/|x|^{r_1}=C$.[16]

14 See Lemma A4 in the supplemental appendix for technical details about this approximation.

15 Moreover, in the current setting, feasible inference requires not only the consistency of the HAC estimator, but also a characterization of its rate of convergence (see Theorem 2(b)).

16 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.

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 (3.1)). More precisely, we maintain the following assumption.

Assumption 7. We have Assumption 5. Moreover, (i) for any $n > 0$, any t and any j, $\mathbb{E}\left[X_{n,t}\right] = 0$ and $\mathbb{E}\left[X_{n,t}X_{n,t+j}\right]$ 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 $X_{n,t}X_{n,t+j}^{\top}$.[17] Conditions (iii) reflect 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 5. *Under Assumptions 6 and 7,* $\|\widetilde{\Sigma}_n - \Sigma_n\| = O_p(\bar{c}_n^2 m_n(M_n k_n^{-1/2} + M_n^{-r_1\wedge r_2}))$.

Comment. Theorem 5 provides an upper bound for the convergence rate of the HAC estimator. It is interesting to note that, in the conventional setting with fixed m_n and $\bar{c}_n = O(1)$, 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 described in Section 2.3 involve estimating/calibrating a finite-dimensional parameter in the structural model. In nonparametric series estimation problems, the HAC estimator is constructed using residuals from the nonparametric regression. We now proceed to extend Theorem 5 to accommodate generated variables.

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

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

where Z_t is observed and $g(z, \theta)$ is a measurable function known up to a parameter 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 (3.5) and (3.6) are then given by

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

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

Assumption 8(i) is a high-level condition that embodies two types of regularities: the smoothness of $g(\cdot)$ with respect to 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 8(ii) states that the m_n-dimensional vector is of size $O(m_n^{1/2})$ in L_2-norm, which holds trivially in most applications.

Theorem 6. *Under Assumptions 6–8 , we have*

$$\left\|\widehat{\Sigma}_n - \Sigma_n\right\| = O_p\left(\bar{c}_n^2 m_n(M_n k_n^{-1/2} + M_n^{-r_1\wedge r_2})\right) + O_p\left(M_n m_n^{1/2}\delta_{\theta,n}\right). \tag{3.7}$$

[17] 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.

Comments. (i) The estimation error shown in (3.7) 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 (3.7) 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 (3.7) is dominated by the first. Simply put, the "plug-in" error resulted from using a parametric preliminary estimator $\hat{\theta}_n$ is negligible compared to the intrinsic sampling variability that is present even in the infeasible case with known 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.

Finally, as an illustration of the use of Theorem 6, we revisit the uniform inference theory formalized by Theorem 2. As discussed in Section 3.1, when (u_t) does not form a martingale difference sequence, $A_n \equiv Var(n^{-1/2}\sum_{t=1}^{n} P(X_t)u_t)$ is generally a long-run covariance matrix. We can estimate A_n using the HAC estimator described above. More specifically, we set

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

and

$$\widehat{A}_n \equiv n^{-1}\sum_{s=-n+1}^{n-1}\mathcal{K}(s/M_n)\,\widehat{\Gamma}_n(s). \tag{3.8}$$

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

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; see, for example, Shimer (2004), Hall (2005), Hall and Milgrom (2008), Mortensen and Nagypál (2017), Gertler and Trigari (2009), and Pissarides (2009), among others.

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

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

We aim to shed some light on this debate from an econometric point of view. Rather than resorting to some "external" calibration target, we rely on a conditional moment restriction that arises "internally" from the equilibrium Bellman equations. Specifically, we apply the proposed nonparametric test as described in Section 2.3, in order to examine whether the calibrated parameters are compatible with the equilibrium conditional moment restriction.

Turning to the details, we first briefly restate HM's version of the standard Mortensen–Pissarides model with aggregate uncertainty. Time is discrete. There is a unit measure of infinitely lived workers and a continuum of infinitely lived firms. The workers maximize their expected lifetime utility and the firms maximize their expected profit. Workers and firms share the same discount factor 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.

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

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

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

$$J_p = p - w_p + \delta\left(1-s\right)\mathbb{E}_p\left[J_{p'}\right] \tag{4.1}$$

$$V_p = -c_p + \delta q\left(\theta_p\right)\mathbb{E}_p\left[J_{p'}\right] \tag{4.2}$$

$$U_p = z + \delta\left\{f\left(\theta_p\right)\mathbb{E}_p\left[W_{p'}\right] + (1 - f\left(\theta_p\right)\mathbb{E}_p\left[U_{p'}\right])\right\} \tag{4.3}$$

$$W_p = w_p + \delta\left\{(1-s)\mathbb{E}_p\left[W_{p'}\right] + s\mathbb{E}_p\left[U_{p'}\right]\right\}. \tag{4.4}$$

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

From these equilibrium conditions, we can solve the functions J_p, V_p, U_p, W_p and w_p in terms of θ_p, and then reduce the system into one functional equation.[18] Instead of solving the fixed-point problem, we replace p and θ with their observed time series, yielding the following the equilibrium conditional moment restriction

$$\mathbb{E}\left[\zeta_{t+1} - z | p_t\right] = 0, \tag{4.5}$$

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

$$\zeta_{t+1} \equiv p_{t+1} - \frac{\beta\theta_{t+1}c_{t+1}}{1-\beta} + \frac{(1-s)\,c_{t+1}}{(1-\beta)\,q\left(\theta_{t+1}\right)} - \frac{c_t}{(1-\beta)\,\delta q\left(\theta_t\right)}. \tag{4.6}$$

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

4.2. Empirical results

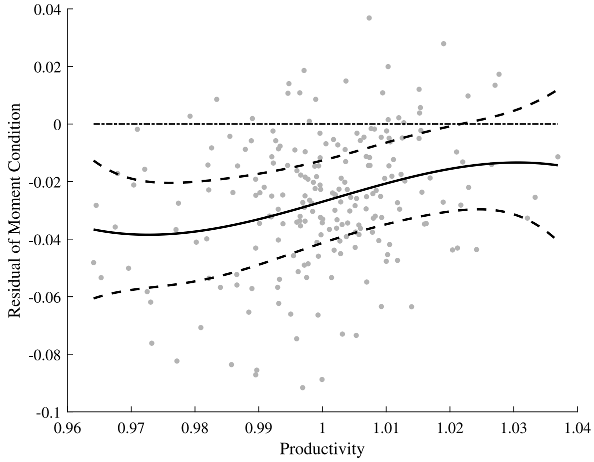
We start with testing whether the equilibrium conditional moment restriction (4.5) holds or not for the parameters calibrated by HM.[19] It is instructive to briefly recall their calibration strategy, which involves two stages. All parameters except for (z, β, l) are calibrated by matching certain empirical quantities in the first stage. The second stage further pins down these three parameters by matching model-implied wage-productivity elasticity, average job finding rate, and average market tightness with their empirical estimates, which is the more contentious part of the calibration (Hornstein et al., 2005). For this reason, we focus on these key parameters so as to directly speak to the core of the unemployment volatility puzzle. The value of nonmarket activity z is of particular importance because it is the sole determinant of the fundamental surplus fraction in the standard Mortensen–Pissarides model (Ljungqvist and Sargent, 2017). We use the same data from 1951 to 2004 as in Hagedorn and Manovskii (2008).[20]

Fig. 1(a) shows the scatter of the residual $\zeta_{t+1} - z$ in the moment condition (4.5) versus the conditioning variable p_t. Under the equilibrium conditional moment restriction, $\zeta_{t+1} - z$ should be centered around zero conditional on each level of p_t, and there should be no correlation pattern between these variables. In contrast, we find that the scatter of $\zeta_{t+1} - z$ is centered below zero, suggesting that z is too high given the other calibrated parameters. In addition, there appears to be a mild positive relationship between the residual and productivity. These patterns are more clearly revealed by the nonparametric fit of $\mathbb{E}[\zeta_{t+1} - z|p_t]$, which is estimated based a cubic polynomial (corresponding to $m_n = 4$) and displayed

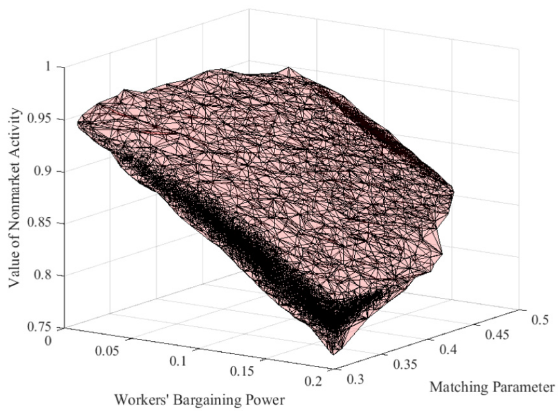
18 The detailed derivation is given in Supplemental Appendix S.A.6.

19 The calibrated parameters play the role of $\hat{\gamma}_n$ in the setting of Section 2.3.

20 The data is obtained from the publisher's website. The p_t and θ_t variables are measured using their cyclical component obtained from the Hodrick–Prescott filter with smoothing parameter 1600. The calibrated parameters are adjusted to the quarterly frequency. We refer the reader to Hagedorn and Manovskii (2008) for additional information about their data and calibration.



(a) Estimated Conditional Moment



(b) Anderson–Rubin Confidence Set

Fig. 1. Nonparametric test for equilibrium conditional moment restriction. Panel (a) shows the scatter of the residual of the equilibrium conditional moment restriction $\zeta_{t+1} - z$ versus the productivity p_t, the nonparametric fit (solid) and the 95% uniform two-sided confidence band (dashed). The series estimator (solid) is computed using a cubic polynomial and the standard error is computed under the martingale difference assumption implied by the conditional moment restriction. Panel (b) plots the 95% Anderson–Rubin confidence set for value of nonmarket activity z, worker's bargaining power β, and matching parameter l.

as the solid line.[21] The uniform confidence band of the conditional moment function does not cover zero for a wide range of productivity levels, indicating a strong rejection (with the p-value being virtually zero) of the equilibrium conditional moment restriction given the calibrated parameter values.

We next ask a more constructive question: Which parameter values, if there are any, are compatible with the equilibrium conditional moment restriction (4.5)? To answer this question formally, we construct the Anderson–Rubin confidence set for (z, β, l) obtained by inverting the nonparametric specification test, while fixing the other parameters at their calibrated values.[22] Fig. 1(b) shows the 3-dimensional 95%-level confidence set. The confidence set is far away from empty, suggesting that the equilibrium conditional moment restriction is compatible with the data for a wide range of parameter values and, to this extent, is not overly restrictive. But we also see that "admissible" values of z is generally notably lower than the calibrated value 0.955. By the theory of Ljungqvist and Sargent (2017), a mild decrease in z can significantly reduce the fundamental surplus fraction. For example, changing z from 0.955 to 0.9 will reduce the fundamental surplus fraction by more than half, and hence causes the unemployment volatility to drop by a similarly amount. Our finding thus suggests that the unemployment volatility puzzle remains a puzzle for the standard model once we insist that the parameters – particularly the value of nonmarket activity – are econometrically compatible with the equilibrium conditional moment restriction.

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. The uniform inference theory relies crucially on our novel strong approximation theory for heterogeneous dependent data with growing dimensions. To conduct feasible inference, we also develop a HAC estimation theory in a setting with increasing dimension. 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.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2019.09.011.

21 The procedure can be implemented using a Stata command "tssreg" that is available at the authors' research websites. In results not presented here, we use Legendre polynomials up to the 8th order (corresponding to $m_n = 9$) and obtain very similar empirical findings.

22 The inversion is implemented by using a (standard) grid search: We consider $z \in [0.01, 0.99]$, $\beta \in [0.01, 0.2]$, $l \in [0.3, 0.5]$, and discretize these intervals with mesh size 0.001.

References

Aït-Sahalia, Y., Jacod, J., 2014. High-Frequency Financial Econometrics. Princeton University Press.
Andrews, D.W.K., 1984. Non-strong mixing autoregressive processes. J. Appl. Probab. 21 (4), 930–934.
Andrews, D.W.K., 1985. A nearly independent, but non-strong mixing, triangular array. J. Appl. Probab. 22 (3), 729–731.
Andrews, D.W.K., 1991a. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. Econometrica 59 (2), 307–345.
Andrews, D.W.K., 1991b. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.
Belloni, A., Chernozhukov, V., Chetverikov, D., Kato, K., 2015. Some new asymptotic theory for least squares series: Pointwise and uniform results. J. Econometrics 186 (2), 345–366, High Dimensional Problems in Econometrics.
Bierens, H., 1982. Consistent model specification tests. J. Econometrics 20, 105–134.
Bierens, H.J., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65 (5), pp. 1129–1151.
Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics, Vol. 6B, first ed. Elsevier.
Chen, X., Christensen, T.M., 2015. Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. J. Econometrics 188 (2), 447–465.
Chen, X., Christensen, T.M., 2018. Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. Quant. Econ. 9 (1), 39–84.
Chen, X., Liao, Z., 2014. Sieve M inference on irregular parameters. J. Econometrics 182 (1), 70–86.
Chen, X., Liao, Z., Sun, Y., 2014. Sieve inference on possibly misspecified semi-nonparametric time series models. J. Econometrics 178, 639–658.
Chen, X., Pouzo, D., 2012. Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. Econometrica 80 (1), 277–321.
Chen, X., Pouzo, D., 2015. Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. Econometrica 83 (3), 1013–1079.
Chen, X., Shen, X., 1998. Sieve extremum estimates for weakly dependent data. Econometrica 289–314.
Chernick, M.R., 1981. A limit theorem for the maximum of autoregressive processes with uniform marginal distributions. Ann. Probab. 9 (1), 145–149.
Chernozhukov, V., Chetverikov, D., Kato, K., 2013a. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. Ann. Statist. 41 (6), 2786–2819.
Chernozhukov, V., Chetverikov, D., Kato, K., 2013b. Testing Many Moment Inequalities. Discussion paper.
Chernozhukov, V., Chetverikov, D., Kato, K., 2014. Gaussian approximation of suprema of empirical processes. Ann. Statist. 42 (4), 1564–1597.
Chernozhukov, V., Lee, S., Rosen, A.M., 2013c. Intersection bounds: Estimation and inference. Econometrica 81 (2), 667–737.
Chodorow-Reich, G., Karabarbounis, L., 2016. The cyclicality of the opportunity cost of employment. J. Political Econ. 124 (6), 1563–1618.
Costain, J.S., Reiter, M., 2008. Business cycles, unemployment insurance, and the calibration of matching models. J. Econom. Dynam. Control 32, 1120–1155.
Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press.
De Jong, R.M., 2000. A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance matrix estimators. Econometric Theory 16 (2), 262–268.
Gallant, A., Souza, G., 1991. On the asymptotic normality of Fourier flexible form estimates. J. Econometrics 50 (3), 329–353.
van de Geer, S., 1990. Estimating a regression function. Ann. Statist. 18 (2), 907–924.
Gertler, M., Trigari, A., 2009. Unemployment fluctuations with staggered Nash Wage bargaining. J. Political Econ. 117 (1), 38–86.
den Haan, W.J., Ramey, G., Watson, J., 2000. Job destruction and propagation of shocks. Am. Econ. Rev. 90 (3), 482–498.
Hagedorn, M., Manovskii, I., 2008. The cyclical behavior of equilibrium unemployment and vacancies revisited. Am. Econ. Rev. 98 (4), 1692–1706.
Hall, R.E., 2005. Employment fluctuations with equilibrium wage stickiness. Am. Econ. Rev. 95 (1), 50–65.
Hall, P., Heyde, C., 1980. Martingale limit theory and its application. In: Probability and Mathematical Statistics: A Series of Monographs and Textbooks, Academic Press.
Hall, R.E., Milgrom, P.R., 2008. The limited influence of unemployment on the wage bargain. Am. Econ. Rev. 98 (4), 1653–1674.
Hannan, E.J., 1970. Multiple Time Series. Wiley, New York.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.
Hansen, B.E., 1992. Consistent covariance matrix estimation for dependent heterogeneous processes. Econometrica 60 (4), 967–972.
Hansen, B., 2015. A Unified Asymptotic Distribution Theory for Parametric and Non- Parametric Least Squares. Working paper.
Hardle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Ann. Statist. 21 (4), 1926–1947.
He, X., Shao, Q., 2000. On parameters of increasing dimensions. J. Multivariate Anal. 73 (1), 120–135.
Hornstein, A., Krusell, P., Violante, G., 2005. Unemployment and vacancy flucutations in the maching model: Inspecting the mechanism. Fed. Res. Bank Richmond Econ. Q. 91 (3), 19–51.
Ibragimov, I., Linnik, Y.V., 1971. Independent and Stationary Sequences of Random Variables. Wolters–Nordhoff, Groningen.
Jacod, J., Protter, P., 2012. Discretization of Processes. Springer.
Jacod, J., Shiryaev, A.N., 2003. Limit Theorems for Stochastic Processes, second ed. Springer-Verlag, New York.
de Jong, R.M., Davidson, J., 2000. Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. Econometrica 68 (2), 407–423.
Kydland, F.E., Prescott, E.C., 1996. The computational experiment: An econometric tool. J. Econ. Perspect. 107 (9), 2630–2665.
Ljungqvist, L., Sargent, T.J., 2017. The fundamental surplus. Amer. Econ. Rev. 107 (9), 2630–2665.
Mortensen, D.T., Nagypál, E., 2017. More on unemployment and vacancy fluctuations. Rev. Econ. Dyn. 107 (9), 2630–2665.
Mortensen, D.T., Pissarides, C.A., 1994. Job creation and job destruction in the theory of unemployment. Rev. Econom. Stud. 61 (3), 397–415.
Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. J. Econometrics 79 (1), 147–168.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.
Pissarides, C.A., 1985. Short-run equilibrium dynamics of unemployment, vacancies, and real wages. Am. Econ. Rev. 75 (4), 676–690.
Pissarides, C., 2000. Equilibrium Unemployment Theory. MIT Press, Cambridge, MA.
Pissarides, C.A., 2009. The unemployment volatility puzzle: Is wage stickiness the answer? Econometrica 77 (5), 1339–1369.
Pötscher, B.M., Prucha, I.R., 2013. Dynamic Nonlinear Econometric Models: Asymptotic Theory. Springer Science & Business Media.
Shimer, R., 2004. The consequences of rigid wages in search models. J. Eur. Econom. Assoc. 2 (2/3), 469–479.
Shimer, R., 2005. The cyclical behavior of equilibrium unemployment and vacancies. Am. Econ. Rev. 95 (1), 25–49.
van der Vaart, A., Wellner, J., 1996. Weak Convergence and Empirical Processes. Springer-Verlag.
White, H., 2001. Asymptotic Theory for Econometricians. Academic Press.
White, H., Gallant, R., 1988. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Oxford: Basil Blackwell.
Wu, W.B., 2005. Nonlinear system theory: Another look at dependence. Proc. Natl. Acad. Sci. 102 (40), 14150–14154.
Yurinskii, V.V., 1978. On the error of the Gaussian approximation for convolutions. Theory Probab. Appl. 22 (2), 236–247.
Zhang, X., Cheng, G., 2018. GaussIan approximation for high dimensional vector under physical dependence. Bernoulli 24 (4A), 2640–2675.
Zhang, D., Wu, W.B., 2017. Gaussian approximation for high dimensional time series. Ann. Statist. 45 (5), 1895–1919.