Uniform Nonparametric Inference for Spatially Dependent Panel Data

Jia Li*  Zhipeng Liao†  Wenyu Zhou‡

Abstract
This paper proposes a uniform functional inference method for nonparametric regressions in a panel-data setting that features general unknown forms of spatio-temporal dependence. The method requires a long time span, but does not impose any restriction on the size of the cross section or the strength of spatial correlation. The uniform inference is justified via a new growing-dimensional Gaussian coupling theory for spatio-temporally dependent panels. We apply the method to study the nonparametric relationship between forecast error and forecast revision of professional forecasters, and document nonparametric evidence for information rigidity, particularly in the presence of high financial risk and uncertainty.

Keywords: coupling, information rigidity, series estimation, spatial dependence, uniform confidence band.

JEL Codes: C14, C22, C32.

*School of Economics, Singapore Management University, Singapore; e-mail: jiali@smu.edu.sg.
†Department of Economics, UCLA, Los Angeles, CA 90095; e-mail: zhipeng.liao@econ.ucla.edu.
‡International Business School, Zhejiang University, Haining, Zhejiang 314400, China; e-mail: wenyuzhou@intl.zju.edu.cn.
1 Introduction

Nonparametric regressions allow empirical researchers to study the conditional mean function of a dependent variable given certain covariates in a flexible manner. While classical methods were originally motivated to study i.i.d. data (Nadaraya (1964), Watson (1964), Andrews (1991a), Newey (1997)), a vast literature has emerged to accommodate both time-series and spatial dependence (Robinson (1983), Robinson (2011), Chen and Shen (1998), Jenish (2012), Chen, Liao, and Sun (2014), Lee and Robinson (2016)). The prior literature has mainly focused on the pointwise inference of the unknown function by providing confidence intervals for the function’s value evaluated at a given point. This may be unsatisfactory in empirical work, because a practitioner’s main goal of performing a nonparametric estimation in the first place is often to make inferential statements regarding the entire function, which would require a uniform inference theory. The contribution of this paper is to develop such a method for panel data, which accommodates general unknown forms of dependence in both time-series and cross-sectional (i.e., spatial) dimensions that are now well known to be important in various areas of economics and finance (Bertrand, Duflo, and Mullainathan (2004), Petersen (2009)).

The key challenge for conducting uniform inference is that the asymptotic analysis for the nonparametric functional estimator is a non-Donsker problem, because the estimator does not admit a functional central limit theorem in the usual weak-convergence sense. This issue is particularly easy to understand in the context of series regression (Andrews (1991a), Newey (1997)), where the nonparametric estimation is carried out by regressing the dependent variable on an asymptotically growing number of approximating functions (e.g., polynomials) of the covariates. Because the dimensionality of the set of regressors increases with the sample size, conventional central limit theorems and the “textbook” notion of convergence in distribution can no longer be used to capture the joint asymptotic normality of the regression coefficients.
In a cross-sectional setting with independent data, Belloni, Chernozhukov, Chetverikov, and Kato (2015) make an important contribution to address this non-Donsker issue. These authors show that the growing-dimensional regression coefficient in the nonparametric series estimation may be strongly approximated, or “coupled,” by a Gaussian random vector. Consequently, the estimation error of the functional estimator may be further strongly approximated by a divergent sequence of Gaussian processes. Intuitively, these Gaussian coupling results are analogous to the usual convergence in distribution for formalizing the notion of asymptotic normality, but the former are valid even in the growing-dimensional context. Belloni et al.’s theory relies on Yurinskii’s coupling for the sample average of independent random vectors and so its application is limited to random samples. Li and Liao (2020) establish a coupling result for general time-series data modeled as heterogeneous mixingales and use it to construct a uniform nonparametric inference procedure for macroeconomic and financial applications.

The use of nonparametric methods in the time-series context, however, may be hindered by a small sample size: For example, the number of observations for macroeconomic time series is typically in the low hundreds. The limited information embodied in the small sample may render nonparametric estimators too noisy to provide interesting empirical discoveries. Panel data is helpful in this regard: If the researcher is willing to assume that the conditional mean function is shared among cross-sectional units (e.g., countries, states, cities, or firms), more accurate nonparametric estimates may be obtained by further pooling the cross-sectional information.

This consideration motivates us to develop a uniform nonparametric inference method tailored for panel-data applications. An interesting aspect of building such a procedure is to accommodate different types of spatio-temporal dependence encountered by empirical researchers (Bertrand, Duflo, and Mullainathan (2004), Petersen (2009)). In the baseline context of linear panel regressions, this mainly manifests as alternative ways of comput-
ing standard errors. In applied microeconomic settings, a popular choice is the clustered standard error proposed by Arellano (1987), which is White’s (1980) standard error formed using the cross section of time-series averages. This approach allows for general serial correlation, but relies on cross-sectional/spatial independence. Ruling out spatial correlation is undesirable for applications in macroeconomics and finance. Driscoll and Kraay (1998) propose an alternative approach under which the standard errors are computed using heteroskedasticity and autocorrelation consistent (HAC) estimators (Newey and West (1987), Andrews (1991a)) of cross-sectional averages. An advantage of the Driscoll–Kraay approach is that it allows for arbitrary spatial dependence in the cross-sectional dimension and, at the same time, it also accommodates a type of “weak” serial dependence commonly employed in time-series analysis.

We develop our uniform nonparametric inference under a similar spatio-temporal dependence structure as in Driscoll and Kraay (1998). Like these authors, we also allow for arbitrary spatial dependence and derive asymptotics under a “large T” setting by exploiting the weak dependence in the time-series dimension. Needless to say, our econometric objective is quite distinct from that prior work: Driscoll and Kraay’s (1998) study is about how to construct a HAC estimator for the standard error of a classical GMM estimator, but we focus on how to make uniform (functional) inference for the conditional expectation function. The key component of our analysis is to establish a growing-dimensional Gaussian coupling theory so as to capture the joint asymptotic normality of the “many” regression coefficients in a nonparametric series regression, which is new to the econometrics literature.

As an empirical illustration of the proposed method, we study the nonparametric relationship between the average ex post forecast error and the ex ante forecast revision in the Survey of Professional Forecasters. In an important paper, Coibion and Gorodnichenko (2015) show that these variables are positively related in the presence of information rigid-
ity (Mankiw and Reis (2002), Sims (2003), Woodford (2003), Reis (2006)). Consistent with this theoretical prediction, our nonparametric estimate of the conditional mean function of forecast error is increasing in the forecast revision. We also find that the linear specification employed by Coibion and Gorodnichenko (2015) is compatible with the observed data based on a formal nonparametric specification test. In addition, we estimate a semi-nonparametric model under which information rigidity is a nonparametric function of a measure of aggregate risk or uncertainty. We find that information rigidity is most pronounced during periods with high stock market volatility and/or financial uncertainty.

The present paper is related to several strands of literature. As mentioned above, the most closely related work includes Driscoll and Kraay (1998), Belloni, Chernozhukov, Chetverikov, and Kato (2015), and Li and Liao (2020), in that we extend the existing uniform inference method from a purely cross-sectional or time-series setting to a more general panel-data setting so as to accommodate Driscoll and Kraay’s dependence structure. The proposed uniform inference theory may be further related to prior work on pointwise inference for independent and serially dependent data; see Robinson (1983), Andrews (1991a), Newey (1997), and Chen, Liao, and Sun (2014), among others.

Like Driscoll and Kraay (1998), we allow the spatial dependence among cross-sectional units to be arbitrarily strong and the size of the cross section may be fixed or grow to infinity. We also exploit the weak dependence in the time-series dimension by using a “large T” asymptotic argument. The proposed method thus cannot be applied to “short” panels that exhibit spatial dependence. Under this alternative setting, Conley (1999) proposes a feasible inference theory for GMM using limit theorems for spatially weakly dependent data (also see Kelejian and Prucha (2007), Kim and Sun (2011)), provided that the spatial dependence among unobservables can be approximately gauged by an observed measure of “economic distance.” In a similar spirit, Robinson (2011), Jenish (2012), and Lee and Robinson (2016) develop results for pointwise nonparametric inference, but the
more challenging uniform inference problem in this context remains to be an open question. It might be interesting to address this issue by developing a growing-dimensional Gaussian coupling theory for spatially weakly dependent random fields. This is beyond the scope of the present paper and is left for future research.

The remainder of the paper is organized as follows. Section 2 presents our uniform nonparametric inference method. Section 3 reports the finite-sample performance of the proposed method in a Monte Carlo study. An empirical illustration is provided in Section 4. Section 5 concludes. All proofs are in the Appendix. The Online Supplemental Appendix contains a battery of robustness checks for our empirical results.

2 The econometric method

In this section, we present the uniform nonparametric inference method. We describe the setting and some relevant background in Section 2.1. Section 2.2 presents growing-dimensional Gaussian coupling results for spatio-temporally dependent panel data, which are then used to construct uniform confidence bands in Section 2.3.

2.1 The setting and background

Consider an \( N \times T \) panel \((Y_{it}, X_{it})_{1 \leq i \leq N, 1 \leq t \leq T}\) where \(Y_{it}\) is a scalar-valued dependent variable and the covariate \(X_{it}\) takes value in a compact set \(\mathcal{X} \subseteq \mathbb{R}^d\). Like Driscoll and Kraay (1998), we are interested in a setting with “weak” time-series dependence, whereas the spatial dependence among cross-sectional units may be arbitrarily strong with an unknown form. Correspondingly, we derive asymptotic results in a “large \( T \)” thought experiment, but do not make any assumption on the cross-sectional dimension \(N\). That is, \(T \to \infty\) and \(N\) may be fixed or grow to infinity.

The inferential target is the conditional expectation function of \(Y_{it}\) given \(X_{it}\), denoted
by \( g(x) \equiv \mathbb{E}[Y_{it}|X_{it} = x] \) for \( x \in \mathcal{X} \). Setting the disturbance term as \( \epsilon_{it} \equiv Y_{it} - g(X_{it}) \), we may equivalently state the problem as a nonparametric regression

\[
Y_{it} = g(X_{it}) + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it}|X_{it}] = 0.
\] (2.1)

Our main goal is to nonparametrically estimate \( g(\cdot) \) and construct a uniform confidence band for it. More precisely, for a given confidence level \( 1 - \alpha \), we aim to construct a pair of functional estimates \([L(\cdot), U(\cdot)]\) such that

\[
P( L(x) \leq g(x) \leq U(x) \text{ for all } x \in \mathcal{X} ) \to 1 - \alpha, \quad \text{as} \quad T \to \infty.
\] (2.2)

Our procedure is built on the series regression method (Andrews (1991a), Newey (1997)). Under the series approach, the nonparametric estimation can be easily performed by running a (pooled) least-squares regression of \( Y_{it} \) on a collection of approximating functions of \( X_{it} \). Specifically, consider a column vector of approximating functions \( P(\cdot) = (p_j(\cdot))_{1 \leq j \leq m} \), which may be polynomials, splines, trigonometric functions, wavelets, etc.; see Chen (2007) for a comprehensive review. Regressing \( Y_{it} \) on \( P(X_{it}) \) yields the regression coefficient

\[
\hat{b} \equiv \left( \sum_{t=1}^{T} \sum_{i=1}^{N} P(X_{it}) P(X_{it})^\top \right)^{-1} \left( \sum_{t=1}^{T} \sum_{i=1}^{N} P(X_{it}) Y_{it} \right),
\] (2.3)

and the resulting nonparametric estimator for \( g(\cdot) \) is then given by

\[
\hat{g}(\cdot) \equiv P(\cdot)^\top \hat{b}.
\] (2.4)

This nonparametric series estimator is very simple to implement and naturally generalizes the commonly used ordinary least-squares. The key element of the nonparametric theory
is to let the number of series terms $m \to \infty$, so that the unknown function $g(\cdot)$ can be well approximated by a growing set of approximating functions. The growing dimensionality is exactly the main source of complication for the theoretical analysis.

The pointwise inference for $g(x)$ at a given point $x \in \mathcal{X}$ has been extensively studied in the prior literature using standard econometric techniques; see, for example, Andrews (1991a) and Newey (1997). The uniform inference, however, is much more challenging because it is a non-Donsker problem (i.e., $\hat{g}(\cdot)$ does not admit a functional central limit theorem in the sense of weak convergence). This theoretical difficulty stems from the growing dimensionality of the series regression. In particular, one cannot use classical central limit theorems to characterize the asymptotic normality of $\hat{b}$, because its dimensionality is divergent asymptotically. This in turn leads to difficulties for establishing the asymptotic Gaussianity of the functional estimator $\hat{g}(\cdot)$.

In a cross-sectional setting (i.e., $T = 1$) with independent data, Belloni, Chernozhukov, Chetverikov, and Kato (2015) show that the aforementioned non-Donsker issue may be addressed by using a strong Gaussian approximation theory. For ease of discussion, denote $h_{it} \equiv P(X_{it})\epsilon_{it}$, so that the score vector for the (single cross-section) series regression may be written as $N^{-1/2} \sum_{i=1}^{N} h_{i1}$. Under the assumption that the cross-sectional units are independent, Belloni et al. invoke Yurinskii’s coupling to show that the growing-dimensional (i.e., $m \to \infty$) score vector may be strongly approximated, or “coupled,” by a zero-mean Gaussian random vector $\xi_N$ with the same variance-covariance matrix, that is,

$$
\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} h_{i1} - \xi_N \right\| = o_p(1), \quad \xi_N \sim \mathcal{N}\left(0, \frac{1}{N} \sum_{i=1}^{N} \text{Var}[h_{i1}] \right),$$

(2.5)

where $\| \cdot \|$ denotes the Euclidean norm. Consequently, the estimation error in $\hat{b}$ also admits
a Gaussian coupling in the form of

\[ \| N^{1/2} \left( \hat{b} - b^* \right) - Q^{-1} \xi_N \| = o_P(1), \]

where \( b^* \) is the population regression coefficient and \( Q \) is the population Gram matrix \( \mathbb{E} [ N^{-1} \sum_{i=1}^{N} P (X_{it}) P (X_{it})^T ] \). The coupling of the regression coefficient in turn implies that the scaled estimation error function \( N^{1/2} (\hat{g} (\cdot) - g (\cdot)) \) may be coupled by a Gaussian process \( P (\cdot)^T Q^{-1} \xi_N \), which can then be used to construct uniform confidence bands for \( g (\cdot) \).

Note that the key to establish a uniform inference theory is the growing-dimensional Gaussian coupling for the score vector as described in (2.5). Li and Liao (2020) generalize Yurinskii’s coupling from the independent-data setting to one with heterogeneous mixingales, which enables them to extend Belloni et al.’s method to various time-series settings. Li and Liao’s coupling theory implies that the score for the \( i \)th time series admits a Gaussian coupling in the form of

\[ \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} h_{it} - \xi_{T}^{(i)} \right\| = o_P(1), \quad \xi_{T}^{(i)} \sim \mathcal{N} \left( 0, \text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} h_{it} \right] \right). \]  

Not surprisingly, the variance-covariance matrix of the coupling variable \( \xi_{T}^{(i)} \) is generally the long-run variance-covariance matrix of the \( (h_{it})_{t \geq 1} \) series. For the conduct of feasible inference, these authors also show that standard HAC estimators (Newey and West (1987), Andrews (1991b)) remain to be valid even under the growing-dimensional setting.

The present paper further extends Li and Liao’s (2020) theory to the panel-data setting. In particular, we aim to do so without restricting (i) the size of the cross section (i.e., \( N \) may be fixed or growing) or (ii) the degree of spatial dependence between cross-sectional units. These features are important in many applied scenarios. A seemingly
natural approach is to “stack” the cross-sectional units into a multivariate time series and then directly apply Li and Liao’s (2020) coupling theory to obtain a “stacked” version of (2.6). This approach, however, would have two drawbacks. Firstly, note that the stacking would substantially increase the dimensionality of the coupling problem and, as a consequence, the joint coupling can only be obtained under very stringent restrictions on how fast $N$ and/or $m$ may grow as $T \to \infty$. As a matter of fact, $N$ could only grow at a much slower rate than $T$, which is undesirable in applications with even moderately large cross sections. Secondly, to conduct feasible inference, one would need to perform a HAC estimation for the $(Nm) \times (Nm)$ long-run variance-covariance matrix of the stacked score vector. A satisfactory HAC estimation is known to be difficult even if $N$ is moderately large and $m$ is fixed (Driscoll and Kraay (1998)). This issue ought to be more severe in the present growing-dimensional setting with $m \to \infty$.

We thus consider an alternative approach that is inspired by Driscoll and Kraay (1998). These authors’ key insight, when applied to the present context, is to rewrite the scaled score vector as

\[
\frac{1}{N\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{N} h_{it} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} H_{t}, \quad \text{where} \quad H_{t} \equiv \frac{1}{N} \sum_{i=1}^{N} h_{it}. \tag{2.7}
\]

This simple rewriting highlights the fact that the analysis for spatially dependent panels closely resembles the (seemingly) simpler time-series problem, except that the $H_{t}$ time series is now “generated” as a cross-sectional average of the unit-specific influence function $h_{it}$. Guided by this powerful idea, we adapt Li and Liao’s (2020) time-series coupling theory to the more general panel-data setting and use it to construct a uniform inference procedure for the conditional expectation function $g(\cdot)$. We now turn to the details.
2.2 Growing-dimensional Gaussian coupling for panel data

We now present the aforementioned new results concerning the growing-dimensional Gaussian coupling for spatio-temporally dependent panel data. This subsection may be skipped by readers who are mainly interested in applications.

The formal theoretical setting is as follows. Let \( h_{it} \) be an \( m \)-dimensional random vector for \( 1 \leq i \leq N \) and \( 1 \leq t \leq T \). For clarity, we introduce a single index \( n \) for the asymptotic stage. As \( n \to \infty \), we have \( T \to \infty \) and \( m \to \infty \), whereas \( N \) may be fixed or \( N \to \infty \). In this subsection, we emphasize the divergent nature of \( T \) and \( m \) by writing \( T_n \) and \( m_n \), respectively. We also consider a filtration \( \mathcal{F}_t \) of information sets. We do not always assume that \( h_{it} \) is measurable with respect to \( \mathcal{F}_t \), but the filtration is useful to specify the serial dependence of \( h_{it} \).

Our goal is to construct a Gaussian coupling for a sequence of \( m_n \)-dimensional random vectors given by

\[
S_n \equiv a_N^{-1} T_n^{-1/2} \sum_{t=1}^{T_n} \sum_{i=1}^{N} h_{it}.
\]

Like Driscoll and Kraay (1998), we rewrite \( S_n \) as

\[
S_n \equiv \frac{1}{\sqrt{T_n}} \sum_{t=1}^{T_n} H_{n,t}, \quad \text{where} \quad H_{n,t} \equiv \frac{1}{a_N} \sum_{i=1}^{N} h_{it}.
\]

The \( n \) subscript in \( H_{n,t} \) highlights the fact that the latter generally forms a triangular array.

The normalizing sequence \( a_N \) is introduced to ensure that \( a_N^{-1} \sum_{i=1}^{N} h_{it} \) is non-degenerate. For example, one may set \( a_N = N \) if the \( h_{it} \) variables are strongly dependent on the cross section, or set \( a_N = N^{1/2} \) when the cross-sectional dependence is weak (e.g., independence). More generally, it is possible to have \( a_N = N^\gamma \) for some \( \gamma \in (1/2, 1) \) if \( h_{it} \) exhibits some form of spatial weak dependence (Conley (1999), Kelejian and Prucha (2007)). Introducing \( a_N \) helps streamline our theoretical presentation, but the user does not need to know its specific form for implementation (because it is canceled through studentization).

Our coupling theory will be developed in two steps: We first consider the baseline case...
in which each time series \((h_{it})_t \geq 1\) forms a martingale difference sequence (MDS) with respect to the filtration \(\mathcal{F}_t\), and then extend this baseline result to the far more general setting in which \((H_{n,t})_t \geq 1\) forms a mixingale. Singling out the former special case is useful, because the simpler MDS structure quite commonly arises from rational-expectation models, for which the feasible inference does not require a HAC estimation.

Regularity conditions for the MDS setting are collected in the following assumption, where \(\|\cdot\|_S\) denotes the matrix spectral norm and \(H_{n,t}^{(j)}\) denotes the \(j\)th component of \(H_{n,t}\).

**Assumption 1.** Suppose that the panel \((h_{it})_{1 \leq i \leq N, 1 \leq t \leq T}\) satisfies the following: (i) for each \((i, t)\), \(h_{it}\) is \(\mathcal{F}_t\)-measurable and \(E[h_{it}|\mathcal{F}_{t-1}] = 0\); (ii) the eigenvalues \(E[H_{n,t}H_{n,t}^\top]\) are uniformly bounded from above and away from zero by some fixed positive constants; (iii) \(E[|H_{n,t}^{(j)}|^3]\) is bounded uniformly for all \((j, t)\); (iv) uniformly for any sequence \(T'_n\) of integers that satisfies \(T'_n \leq T_n\) and \(T'_n/T_n \to 1\),

\[
\left\| \frac{1}{T_n} \sum_{t=1}^{T'_n} (E[H_{n,t}H_{n,t}^\top|\mathcal{F}_{t-1}] - E[H_{n,t}H_{n,t}^\top]) \right\|_S = O_p(r_n), \quad (2.8)
\]

where \(r_n\) is a real sequence such that \(r_n = o(1)\).

Condition (i) of Assumption 1 states that \(h_{it}\) forms an MDS for each \(i\) with respect to the filtration \(\mathcal{F}_t\), which further implies that \(H_{n,t}\) is a martingale difference array. Condition (ii) requires that the variance-covariance matrix of \(H_{n,t}\) is non-degenerate, and condition (iii) further imposes a bound on its third moment. Condition (iv) mainly requires that the conditional variance-covariance matrix \(E[H_{n,t}H_{n,t}^\top|\mathcal{F}_{t-1}]\) satisfies a matrix law of large numbers at a certain convergence rate. It is worth noting that this condition holds trivially for \(r_n = 0\) if \(H_{n,t}\) is conditionally homoskedastic (i.e., the conditional second moments coincide with the unconditional ones). Overall, Assumption 1 is relatively easy to verify given known results in the literature. It is instructive to consider an example for illustration.
Example 1 (Spatially Dependent Martingale Differences) Suppose that \( h_{it} = f_t + v_{it} \), where the common factor \( f_t \) is an MDS with respect to a filtration \( \mathcal{F}_t \) and, for each \( i \), the unit-specific shock \( v_{it} \) is \( \mathcal{F}_t \)-measurable satisfying \( \mathbb{E}[v_{it} | \mathcal{F}_{t-1}] = 0 \) and \( \mathbb{E}[v_{it}v_{it}^\top | \mathcal{F}_{t-1}] = \Sigma_v \) for some nonrandom matrix \( \Sigma_v \). Suppose that, conditional on \( \mathcal{F}_{t-1} \), \( f_t \) is independent of \( v_{it} \) and the \((v_{it})_{1 \leq i \leq N}\) variables are mutually independent. Under this model, the \( h_{it} \) variables are strongly correlated on the cross section through the common factor \( f_t \). It is easy to see that \( h_{it} \) forms an MDS. To verify conditions (ii) and (iii) in Assumption 1, we set \( a_N = N \) and note that \( H_{n,t} = f_t + \bar{v}_{nt} \) where \( \bar{v}_{nt} \equiv N^{-1} \sum_{i=1}^N v_{it} \). Therefore, by the conditional independence between \( f_t \) and \( v_{it} \), we have \( \mathbb{E}[H_{n,t}H_{n,t}^\top | \mathcal{F}_{t-1}] = \mathbb{E}[f_tf_t^\top | \mathcal{F}_{t-1}] + N^{-1}\Sigma_v \) and \( \mathbb{E}[H_{n,t}H_{n,t}^\top] = \mathbb{E}[f_tf_t^\top] + N^{-1}\Sigma_v \). Condition (ii) is satisfied if the eigenvalues of \( \mathbb{E}[f_tf_t^\top] \) are bounded from above and away from zero, and those of \( \Sigma_v \) are bounded. By Minkowski’s inequality, it is easy to verify condition (iii), provided that the entries of \( f_t \) and \( v_{it} \) have bounded third moments. Finally, note that condition (iv) amounts to

\[
\left\| \frac{1}{T_n} \sum_{t=1}^{T_n} (\mathbb{E}[f_tf_t^\top | \mathcal{F}_{t-1}] - \mathbb{E}[f_tf_t^\top]) \right\|_S = O_p(r_n),
\]

which can be verified by invoking a matrix law of large numbers for the matrix-valued time series \( \mathbb{E}[f_tf_t^\top | \mathcal{F}_{t-1}] \) (see, e.g., Lemma 2.1 in Chen and Christensen (2015) and Proposition B1 in the supplemental appendix of Li and Liao (2020)).

Theorem 1, below, establishes the Gaussian coupling for the \( S_n \) statistic when \( h_{it} \) forms an MDS.

**Theorem 1.** Under Assumption 1, there exists a sequence \( \xi_n \) of \( m_n \)-dimensional random vectors with distribution \( \mathcal{N}(0, T_n^{-1} \sum_{t=1}^{T_n} \text{Var}[H_{n,t}]) \) such that

\[
\|S_n - \xi_n\| = O_p(m_n^{1/2}r_n^{1/2} + T_n^{-1/6}m_n^{5/6}).
\]
A couple of remarks are in order. Firstly, we note that the variance-covariance matrix of the coupling variable is $T_n^{-1} \sum_{t=1}^{T_n} \text{Var}[H_{n,t}]$, which does not involve any autocovariance, because $h_{it}$ forms an MDS with respect to the common filtration $\mathcal{F}_t$. Consequently, the related feasible inference will not require a HAC estimation. Secondly, observe that the rate of convergence of the coupling error is the same as what Li and Liao (2020) obtain in the time-series setting. As alluded above, if we had directly applied Li and Liao’s result by stacking the cross-sectional units into an $N m_n$-dimensional time series, the resulting rate (i.e., $N^{1/2} m_n^{1/2} r_n^{1/2} + T_n^{-1/6} N^{5/6} m_n^{5/6}$) would be much slower when $N$ is large. We have avoided this issue by relying on Driscoll and Kraay’s (1998) insight.

We next extend Theorem 1 to the more general case in which the triangular array $H_{n,t} = a_n^{-1} \sum_{i=1}^{N} h_{it}$ forms a mixingale. The mixingale assumption is stated as follows: For a sequence of constants $\bar{c}_n = O(1)$ and a summable nonnegative sequence $(\psi_k)_{k \geq 0}$ (i.e., $\sum_{k \geq 0} \psi_k < \infty$), we have for $1 \leq j \leq m_n$ and $k \geq 0$,

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

(2.10)

where $\| \cdot \|_q$ denotes the $L_q$-norm of a random variable for some $q \geq 1$, and the constants $\bar{c}_n$ and $\psi_k$ control the magnitude and the serial dependence of the $(H_{n,t})_{t \geq 1}$ variables, respectively. Note that if $H_{n,t}$ forms a martingale difference array and each of its entries has bounded $q$th moment, it is trivially a mixingale that verifies (2.10) with $\psi_k = 0$ for all $k \geq 1$. It is also well known that the mixingale class includes linear processes (e.g., ARMA), various mixing processes, and certain near-epoch processes as special cases, and hence, accommodates a majority of dependence structures seen in time-series econometrics; we refer the reader to Davidson’s (1994) monograph for a comprehensive review of these well-known facts. To derive general results, it is convenient to directly impose a mixingale structure on $H_{n,t}$, instead of specifying conditions on the $h_{it}$ variables. But
the following example illustrates how one may establish (2.10) “from scratch” based on conditions on $h_{it}$.

**Example 2 (Spatially Dependent Mixingales)** Consider the same setting as in Example 1 with one modification: $f_t$ is no longer an MDS, but is a mixingale satisfying, for $1 \leq j \leq m_n$ and $k \geq 0$,

$$
\left\| E[f_t^{(j)}|\mathcal{F}_{t-k}] \right\|_q \leq \bar{c}_{f,n}\psi_{f,k}, \quad \left\| f_t^{(j)} - E[f_t^{(j)}|\mathcal{F}_{t+k}] \right\|_q \leq \bar{c}_{f,n}\psi_{f,k+1}.
$$

Recall from Example 1 that $H_{n,t} = f_t + \bar{v}_{n,t}$ and $\bar{v}_{n,t}$ is a martingale difference array with respect to $\mathcal{F}_t$. Therefore, for $k \geq 0$,

$$
E[H_{n,t}^{(j)}|\mathcal{F}_{t-k}] = E[f_t^{(j)}|\mathcal{F}_{t-k}] + \bar{v}_{n,t}1_{k=0}, \quad H_{n,t}^{(j)} - E[H_{n,t}^{(j)}|\mathcal{F}_{t+k}] = f_t^{(j)} - E[f_t^{(j)}|\mathcal{F}_{t+k}].
$$

Further assume that the $L_q$-norm of each component of $v_{it}$ is bounded by a constant $\bar{c}_v$. By Minkowski’s inequality, we readily verify (2.10) by setting $\bar{c}_n = \bar{c}_{f,n} + \bar{c}_v$, $\psi_0 = \max\{\psi_{f,0}, 1\}$, and $\psi_k = \psi_{f,k}$ for $k \geq 1$.

Although the mixingale concept is far more general than MDS, extending Theorem 1 to the mixingale case is relatively straightforward because the sum of mixingales can be approximated by a martingale, through a technique known as “martingale approximation.” Indeed, under the maintained assumption $\sum_{k\geq 0} \psi_k < \infty$, it can be shown that

$$
\left\| S_n - \frac{1}{\sqrt{T_n}} \sum_{t=1}^{T_n} H_{n,t}^* \right\| = O_p(T_n^{-1/2}m_n^{1/2}), \quad (2.11)
$$

where

$$
H_{n,t}^* = \sum_{s=-\infty}^{\infty} \{E[H_{n,t+s}|\mathcal{F}_t] - E[H_{n,t+s}|\mathcal{F}_{t-1}]\}. \quad (2.12)
$$
Observe that $H_{n,t}^*$ forms a martingale difference array and so $T_n^{-1/2} \sum_{t=1}^{T_n} H_{n,t}^*$ admits a strong Gaussian coupling by Theorem 1. Since $S_n$ can be approximated by $T_n^{-1/2} \sum_{t=1}^{T_n} H_{n,t}^*$ up to a relatively small $O_p(T_n^{-1/2} m_n^{1/2})$ error, this further implies that $S_n$ also admits a Gaussian coupling. Theorem 2, below, formalizes this logic.

**Theorem 2.** Suppose (i) the triangular array $H_{n,t}$ forms a mixingale satisfying (2.10) for some $q \geq 3$; (ii) the martingale difference array $H_{n,t}^*$ defined in (2.12) satisfies Assumption 1; (iii) the largest eigenvalue of $\text{Var}[S_n]$ is bounded; and (iv) $m_n = o(T_n)$. Then there exists a sequence $\xi_n$ of $m_n$-dimensional random vectors with distribution $N(0, \text{Var}[S_n])$ such that

$$\|S_n - \xi_n\| = O_p(m_n^{1/2} t_n^{1/2} + T_n^{-1/6} m_n^{5/6}).$$

(2.13)

Theorem 2 establishes the strong Gaussian approximation for $S_n$ when $H_{n,t}$ forms a mixingale. The convergence rate in (2.13) is the same as that in Theorem 1. It is also important to note that $\text{Var}[S_n]$ is a long-run variance-covariance matrix given by $\text{Var}[S_n] = \text{Var}[T_n^{-1/2} \sum_{t=1}^{T_n} H_{n,t}]$ that generally involves all autocovariances $\text{Cov}[H_{n,t}, H_{n,s}]$. Since $H_{n,t} = a_N^{-1} \sum_{i=1}^{N} h_{it}$, we may further rewrite $\text{Cov}[H_{n,t}, H_{n,s}] = a_N^{-2} \sum_{i,j=1}^{N} \text{Cov}[h_{it}, h_{js}]$, which clarifies how the spatio-temporal correlation across the panel contributes to the sampling variability in $S_n$.

### 2.3 Uniform nonparametric inference procedures for panel data

The Gaussian coupling results developed in the previous subsection (Theorems 1 and 2) allow us to construct uniform confidence bands for the conditional expectation function $g(\cdot)$. The formal theory is straightforward to establish, as it mainly requires verifying the high-level conditions in Theorem 2 of Li and Liao (2020) using the new coupling theorems. For brevity, we omit those technical details, but instead focus on discussing the implementation and the underlying intuition, which may be more useful for guiding
applications.

The heuristic for the uniform inference procedure is as follows. Recall from (2.3) that \( \hat{b} \) is the least-squares coefficient obtained by regressing \( Y_{it} \) on \( P(X_{it}) \), and \( \hat{g}(\cdot) = P(\cdot)^T \hat{b} \) is the nonparametric series estimator for \( g(\cdot) \). When the number of series terms \( m \) is large, we have \( g(\cdot) \approx P(\cdot)^T b^* \) for some “population” regression coefficient \( b^* \) and so \( Y_{it} \approx P(X_{it})^T b^* + \epsilon_{it} \). Hence, with \( \hat{Q} \equiv (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} P(X_{it})P(X_{it})^T \), we have

\[
\hat{b} - b^* \approx \hat{Q}^{-1} \left( \frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} P(X_{it}) \epsilon_{it} \right),
\]

(2.14)

which obviously resembles the representation of the estimation error in the “textbook” least-squares regression, though in the latter case (2.14) would hold as an equality.

The approximation in (2.14) suggests that the asymptotic normality for \( \hat{b} \) may be obtained by applying the coupling theorem for the panel \( h_{it} = P(X_{it}) \epsilon_{it} \). To do so, we set

\[
H_t = a_{-1} \sum_{i=1}^{N} h_{it}
\]

as in Section 2.2 (the \( n \) subscript is omitted here for simplicity). With this notation, we rewrite (2.14) as

\[
\frac{T^{1/2}N}{a_N} (\hat{b} - b^*) \approx \hat{Q}^{-1} \times \frac{1}{\sqrt{T}} \sum_{t=1}^{T} H_t.
\]

(2.15)

Theorem 2 above shows that \( T^{-1/2} \sum_{t=1}^{T} H_t \) may be strongly approximated by a Gaussian vector \( \xi_n \sim N(0, A) \), where

\[
A \equiv \text{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} H_t \right] = \text{Var} \left[ \frac{1}{a_N \sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{N} P(X_{it}) \epsilon_{it} \right].
\]

(2.16)

The estimation error in \( \hat{b} \) thus admits the following Gaussian approximation in distribution:

\[
\frac{T^{1/2}N}{a_N} (\hat{b} - b^*) \overset{d}{\approx} \left( Q^{-1} A Q^{-1} \right)^{1/2} \mathcal{N}_m^*,
\]

(2.17)
where \( Q \equiv (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[P(X_{it})P(X_{it})^\top] \) is the population analogue of \( \hat{Q} \) and \( N_m^* \) is a generic copy of an \( m \)-dimensional standard normal vector. Since \( g(\cdot) \approx P(\cdot)^\top b^\ast \) and \( \hat{g}(\cdot) = P(\cdot)^\top \hat{b} \), we have the following analogous result for the functional estimator

\[
\frac{T^{1/2}}{a_N} (\hat{g}(\cdot) - g(\cdot)) \approx P(\cdot)^\top (Q^{-1} AQ^{-1})^{1/2} N_m^*.
\] (2.18)

In particular, the standard error function for the scaled functional estimation error displayed on the left-hand side is given by

\[
\sigma(\cdot) \equiv \sqrt{P(\cdot)^\top Q^{-1} AQ^{-1} P(\cdot)}.
\] (2.19)

To carry out the feasible inference, we need a consistent estimator for the long-run variance-covariance matrix \( A \). A natural choice is the Newey–West estimator given by

\[
\begin{cases}
\hat{A}(a_N) \equiv \sum_{s=-(M_n-1)}^{M_n-1} \frac{|M_n-s|}{M_n} \hat{\Gamma}_s(a_N), \\
\hat{\Gamma}_s(a_N) \equiv \frac{1}{T} \sum_{t=\max(1,s+1)}^{\min(T-s,T)} \left( \frac{1}{a_N} \sum_{i=1}^{N} P(X_{it}) \hat{\epsilon}_{it} \right) \left( \frac{1}{a_N} \sum_{i=1}^{N} P(X_{it+s}) \hat{\epsilon}_{it+s} \right)^\top,
\end{cases}
\] (2.20)

\( \hat{\epsilon}_{it} = Y_{it} - \hat{g}(X_{it}) \), and \( M_n \) is the bandwidth parameter for nonparametric HAC estimation which may be chosen using Andrews’s (1991b) procedure. If the number of regressors \( m \) were fixed, we could directly use Driscoll and Kraay’s (1998) theory to justify the consistency of \( \hat{A}(a_N) \). However, since \( m \to \infty \) in the present nonparametric setting, we instead invoke the HAC estimation theory for growing-dimensional triangular arrays developed in Li and Liao (2020); see their Theorem 6. In applications, it is useful to note that if \( P(X_{it})\epsilon_{it} \) forms an MDS, then the user does not need to include sample autocovariances in \( \hat{A}(a_N) \), which amounts to setting \( M_n = 1 \). Equipped with \( \hat{A}(a_N) \), we estimate the
standard error function $\sigma (\cdot)$ defined in (2.19) via its sample analogue

$$\hat{\sigma} (\cdot ; a_N) \equiv \sqrt{P (\cdot)^\top \hat{Q}^{-1} \hat{A} (a_N) \hat{Q}^{-1} P (\cdot)}.$$  \hfill (2.21)

For notational simplicity in the discussion below, we omit $a_N$ in the notation when $a_N = N$; in particular, we write $\hat{A} = \hat{A} (N)$ and $\hat{\sigma} (\cdot) = \hat{\sigma} (\cdot ; N)$. It is worth noting that $\hat{A}$ and $\hat{\sigma} (\cdot)$ are feasible estimators as they no longer involve the generally unknown normalizing sequence $a_N$.

Our uniform nonparametric inference is based on a “sup-t” statistic defined as

$$\hat{\tau} \equiv \sup_{x \in \mathcal{X}} \frac{T^{1/2} N}{\sigma (x ; a_N)} \left| \hat{g} (x) - g (x) \right|.$$

At first glance, this statistic might appear “infeasible” because we do not assume that the form of the $a_N$ normalizing factor is known a priori. This is not an issue, because $\hat{\tau}$ is in fact invariant to $a_N$. To see this, we note that $\hat{\sigma} (\cdot ; a_N) = N (a_N^{-1} \hat{\sigma} (\cdot))$ by definition, and hence, we may rewrite $\hat{\tau}$ as $\hat{\tau} \equiv \sup_{x \in \mathcal{X}} T^{1/2} [\hat{g} (x) - g (x)] / \hat{\sigma} (x)$. In view of the Gaussian approximation in (2.18), we may approximate the distribution of $\hat{\tau}$ via that of

$$\sup_{x \in \mathcal{X}} \frac{\left| P (x)^\top (Q^{-1} A Q^{-1})^{1/2} N_m^* \right|}{\sigma (x)},$$

for which a feasible approximation may be further constructed as

$$\hat{\tau}^* \equiv \sup_{x \in \mathcal{X}} \frac{\left| P (x)^\top (Q^{-1} A Q^{-1})^{1/2} N_m^* \right|}{\hat{\sigma} (x ; a_N)} = \sup_{x \in \mathcal{X}} \frac{\left| P (x)^\top (Q^{-1} A Q^{-1})^{1/2} N_m^* \right|}{\hat{\sigma} (x)}.$$

Critical values for the sup-t statistic can be computed as the tail quantile of $\hat{\tau}^*$ via simulation, which in turn can be used to construct the uniform confidence bands for $g (\cdot)$. 

19
For ease of application, we detail the construction of a $1 - \alpha$ level two-sided uniform confidence band for $g(\cdot)$ in the following algorithm.

**Algorithm 1 (Uniform Confidence Band for $g(\cdot)$)**

**Step 1.** Run a pooled panel least-squares regression for $Y_{it}$ on $P(X_{it})$ and obtain $\hat{b}$ as described in (2.3). Set the nonparametric estimator $\hat{g}(\cdot) = P(\cdot)^\top \hat{b}$.

**Step 2.** Let $\hat{\epsilon}_{it} = Y_{it} - \hat{g}(X_{it})$ and $\hat{Q} \equiv (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} P(X_{it})P(X_{it})^\top$. Compute $\hat{A} = \hat{A}(N)$ and $\hat{\sigma}(\cdot) = \hat{\sigma}(\cdot; N)$ according to (2.20) and (2.21) with $a_N = N$.

**Step 3.** Draw $N_m^*$ many times from the $m$-dimensional standard normal distribution. For each draw, compute

$$\hat{\tau}^* = \sup_{x \in \mathcal{X}} \left| P(x)^\top \left( \hat{Q}^{-1} \hat{A} \hat{Q}^{-1} \right)^{1/2} \hat{N}_m^* \right| \hat{\sigma}(x),$$

where the supremum may be computed on a discretized mesh of $\mathcal{X}$. Set the critical value $cv_{1-\alpha}$ as the $1 - \alpha$ empirical quantile of the simulated $\hat{\tau}^*$.

**Step 4.** Report the $1 - \alpha$ level two-sided uniform confidence band for $g(\cdot)$ as $[\hat{g}(\cdot) - cv_{1-\alpha}T^{-1/2}\hat{\sigma}(\cdot), \hat{g}(\cdot) + cv_{1-\alpha}T^{-1/2}\hat{\sigma}(\cdot)]$.

As mentioned above, given the coupling theorems established in Section 2.2, the theoretical validity of the confidence band proposed above can be established by applying Theorem 2 in Li and Liao (2020). We may also adapt this procedure to make uniform inference for the derivative function of $g(\cdot)$. For ease of discussion, we consider the case with scalar-valued $X_{it}$ and denote the derivative of $g(\cdot)$ by $\partial g(\cdot)$.

**Algorithm 2 (Uniform Confidence Band for $\partial g(\cdot)$)**

**Step 1.** Compute $\hat{b}, \hat{Q}$, and $\hat{A}$ as described in Algorithm 1. Set

$$\partial \hat{g}(\cdot) = \partial P(\cdot)^\top \hat{b}, \quad \hat{\sigma}(\cdot) = \sqrt{\partial P(x)^\top \left( \hat{Q}^{-1} \hat{A} \hat{Q}^{-1} \right) \partial P(x)}.$$
Step 2. Draw $N_m^*$ from the $m$-dimensional standard normal distribution many times. For each draw, compute

$$
\hat{\tau}^{i*} = \sup_{x \in \mathcal{X}} \left| \frac{\partial P(x) \top \left( \hat{Q}^{-1} \hat{A} \hat{Q}^{-1} \right)^{1/2} N_m^*}{\tilde{\sigma}(x)} \right|
$$

where the supremum may be computed on a discretized mesh of $\mathcal{X}$. Set the critical value $cv'_{1-\alpha}$ as the $1 - \alpha$ empirical quantile of the simulated $\hat{\tau}^{i*}$.

Step 3. Report the $1 - \alpha$ level two-sided uniform confidence band for $\partial g(\cdot)$ as $[\partial \hat{g}(\cdot) - cv'_{1-\alpha} T^{-1/2} \tilde{\sigma}(\cdot), \partial \hat{g}(\cdot) + cv'_{1-\alpha} T^{-1/2} \tilde{\sigma}(\cdot)]$. $\square$

3 Simulations

We examine the finite-sample performance of the proposed uniform inference method. Section 3.1 describes the setting and Section 3.2 reports the results.

3.1 The setting

We consider the following data generating process (DGP) for the $(Y_{it}, X_{it})$ panel. The underlying nonparametric regression model satisfies

$$
Y_{it} = g(X_{it}) + \epsilon_{it}, \quad \text{where } g(x) = \frac{\exp(x)}{1 + \exp(x)}.
$$

Following Driscoll and Kraay (1998), we simulate $\epsilon_{it}$ from a one-factor model given by

$$
\begin{align*}
\epsilon_{it} &= \sigma_{i} (\lambda_{i} f_{i}^{t} + v_{it}^{\epsilon}), \\
\sigma_{it} &\sim i.i.d. \mathcal{N}(0, 1), \\
\rho &\in \{0, 0.5\}, \\
\end{align*}
$$

$$
\begin{align*}
\epsilon_{it} &= \rho f_{i}^{t-1} + u_{it}^{\epsilon}, \\
\epsilon_{it} &\sim i.i.d. \mathcal{N}(0, 1), \\
\end{align*}
$$

$$
\begin{align*}
\epsilon_{it} &= \rho f_{i}^{t-1} + u_{it}^{\epsilon}, \\
\epsilon_{it} &\sim i.i.d. \mathcal{N}(0, 1), \\
\end{align*}
$$

$\rho \in \{0, 0.5\}$. 

21
where the \( v_{it}^\epsilon \) and \( u_t^\epsilon \) variables are mutually independent in both time-series and cross-sectional dimensions. Under this construction, the factor \( f_t^\epsilon \) has a standard normal distribution with autocorrelation \( \rho \), and \( \epsilon_{it} \sim \mathcal{N}(0, \sigma^2_{\epsilon}) \). The contemporaneous correlation between \( \epsilon_{it} \) and \( \epsilon_{jt} \) is \( \lambda_i \lambda_j \) for \( i \neq j \). We draw the \( \lambda_i \) loadings independently from the Uniform\([0, 1]\) distribution so that the average cross-sectional correlation \( \mathbb{E}[\lambda_i \lambda_j] \) is 0.25.

The \( X_{it} \) variables are simulated in a similar fashion according to

\[
\begin{align*}
X_{it} &= \sigma_X (\lambda_i f_t^X + v_{it}^X), & v_{it}^X \overset{i.i.d.}{\sim} \mathcal{N}(0, 1 - \lambda_i^2), \\
f_t^X &= \rho f_{t-1}^X + u_t^X, & u_t^X \overset{i.i.d.}{\sim} \mathcal{N}(0, 1 - \rho^2),
\end{align*}
\]

where the \((v_{it}^X, u_t^X)\) variables are independent of \((v_{it}^\epsilon, u_t^\epsilon)\). We set the standard deviation of \( X_{it} \) to \( \sigma_X = 2.5 \) so that \( g(\cdot) \) is visibly nonlinear on the \([-2\sigma_X, 2\sigma_X]\) interval. The resulting variance of \( g(X_{it}) \) is approximately 0.12. We further set \( \sigma^2_{\epsilon} = 0.04 \) so that roughly 75% of \( Y_{it} \)'s variance is contributed by \( g(X_{it}) \).

We need a collection of approximating functions to implement the nonparametric series regression. Following Li, Liao, and Quaedvlieg (2020), we first rescale \( X_{it} \) to the \([-1, 1]\) interval using the transformation \( x \mapsto 2\Phi((x - \mu_X)/s_X) - 1 \), where \( \Phi(\cdot) \) is the standard normal distribution function and \( \mu_X \) and \( s_X \) are the sample mean and standard deviation of \( X_{it} \), respectively. The approximating functions are then set to be Legendre polynomials of the transformed \( X_{it} \).\(^1\) Evidently, this type of monotone transformation on \( X_{it} \) is innocuous in the nonparametric estimation context. The main purpose of doing so is to make the regressors approximately orthogonal, which generally improves the numerical stability of the series regression, especially when a large number of series terms are included.\(^2\)

\(^1\)Recall that the \( k \)-th order Legendre polynomial is given by \( \mathcal{L}_k(x) = \frac{1}{2^k k!} \frac{d^k}{dx^k}(x^2 - 1)^k \), and \( \mathcal{L}_j(\cdot) \) is orthogonal to \( \mathcal{L}_k(\cdot) \) under the Lebesgue measure on \([-1, 1]\) for \( j \neq k \).

\(^2\)See Li, Liao, and Gao (2020) for a more detailed discussion on the practical implementation of series estimation.
In each simulation, we compute the proposed uniform confidence band for $g(\cdot)$ at the 95% confidence level as described in Algorithm 1 and report its uniform coverage probability calculated using 10,000 Monte Carlo replications. For comparison, we also report similar results for the confidence bands proposed by Belloni, Chernozhukov, Chetverikov, and Kato (2015) and Li and Liao (2020). Recall that Belloni et al.’s method is valid for independent data, and Li and Liao’s method allows for serial dependence but rules out spatial dependence. For ease of reference, we label these two benchmark methods as BCCK and LL, respectively, and refer to our proposal as the robust method.

We conduct Monte Carlo experiments for a variety of scenarios. Specifically, we consider $T \in \{200, 500, 2000\}$. The small-sample case with $T = 200$ may be relevant for macroeconomic settings (e.g., 50 years of quarterly data), whereas the large sample size with $T = 2000$ is easily obtainable on the daily frequency. Since our asymptotic theory relies on $T \to \infty$, we expect the proposed method to have better performance in the latter setting. To illustrate how the size of the cross section may affect inference in the presence of spatial dependence, we also consider $N \in \{5, 25, 100\}$. Intuitively, since the BCCK and LL methods ignore the (positive) spatial dependence, they tend to underestimate the sampling variability of the functional estimator, which in turn leads to under-coverage, especially when $N$ is large. Finally, we check the robustness of the proposed procedure with respect to the number of series terms by setting $m \in \{6, 8, 10, 12\}$; for example, when $m = 12$, the approximating functions include the constant term and Legendre polynomials up to the eleventh power.

### 3.2 Results

Table 1 reports the uniform coverage rates for the case with $\rho = 0$. In this setting, there is no serial correlation in the data, but spatial dependence is present. Looking at the $N = 5$ case displayed on the top panel, we see that the proposed robust method delivers ade-
Table 1: Coverage Rates of Uniform Confidence Bands ($\rho = 0$)

<table>
<thead>
<tr>
<th></th>
<th>Robust</th>
<th></th>
<th>BCCK</th>
<th></th>
<th></th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200</td>
<td>500</td>
<td>2000</td>
<td>200</td>
<td>500</td>
<td>2000</td>
</tr>
</tbody>
</table>

**Case I: $N = 5$**

| $m = 6$  | 0.923  | 0.937  | 0.943 | 0.883  | 0.893  | 0.892 |
| $m = 8$  | 0.910  | 0.936  | 0.944 | 0.885  | 0.898  | 0.906 |
| $m = 10$ | 0.903  | 0.930  | 0.941 | 0.887  | 0.903  | 0.910 |
| $m = 12$ | 0.894  | 0.922  | 0.942 | 0.885  | 0.899  | 0.914 |

**Case II: $N = 25$**

| $m = 6$  | 0.923  | 0.934  | 0.938 | 0.617  | 0.614  | 0.589 |
| $m = 8$  | 0.920  | 0.935  | 0.946 | 0.672  | 0.658  | 0.666 |
| $m = 10$ | 0.925  | 0.937  | 0.947 | 0.703  | 0.707  | 0.705 |
| $m = 12$ | 0.917  | 0.934  | 0.944 | 0.730  | 0.731  | 0.730 |

**Case III: $N = 100$**

| $m = 6$  | 0.927  | 0.933  | 0.930 | 0.209  | 0.205  | 0.151 |
| $m = 8$  | 0.926  | 0.934  | 0.947 | 0.261  | 0.263  | 0.259 |
| $m = 10$ | 0.922  | 0.940  | 0.947 | 0.293  | 0.298  | 0.291 |
| $m = 12$ | 0.921  | 0.934  | 0.943 | 0.327  | 0.324  | 0.327 |

Note: This table reports the coverage rates of 95% uniform confidence bands for the proposed (spatio-temporal) robust method, as well as the methods of Belloni, Chernozhukov, Chetverikov, and Kato (2015) and Li and Liao (2020), labeled as robust, BCCK, and LL, respectively. The autoregressive coefficient is set to $\rho = 0$. The HAC estimation is performed using the Newey–West estimator with bandwidth $\left\lfloor 0.75T^{1/3} \right\rfloor$. The coverage rates are computed based on 10,000 Monte Carlo replications.
quate finite-sample coverage. In the small-sample case with $T = 200$, we observe 3–5% under-coverage. We also confirm that this is mainly a small-sample issue, because the size distortion essentially disappears when $T = 2000$. These findings also appear to be robust with respect to the number of series terms $m$. On the other hand, the BCCK and LL methods both exhibit more severe size distortions than the robust method. It is important to note that increasing $T$ does little for reducing the under-coverage of these benchmark methods, which suggests that their lack of coverage is not a small-sample issue, but rather because they fail to account for spatial dependence in the data.

The contrast between the robust method and the two benchmarks become even more clear for larger cross sections (i.e., $N = 25$ or 100). We only focus on the $N = 100$ case for brevity. From the bottom panel of Table 1, we see that the coverage rates of the proposed robust method are quite close to the 95% nominal level. In sharp contrast, the coverage rates of the BCCK and LL methods range from 15% to 36%, exhibiting extremely severe under-coverage in this large-$N$ scenario.

We next turn to the case with $\rho = 0.5$, so that both serial and spatial dependence are present in the data. The results are reported in Table 2. Here, we observe that the robust method has more under-coverage than it does in the $\rho = 0$ case, which might be attributed to the well-known difficulty in HAC estimation in the presence of serial dependence. Nevertheless, we still find that the proposed robust method has far better coverage properties than the BCCK and LL methods.

Overall, these simulation results show that the proposed uniform confidence band has adequate finite-sample coverage properties. Its performance appear to be robust with respect to the size of the cross section and the number of approximating functions used in the nonparametric series regression. In contrast, in the presence of spatial dependence, the benchmark BCCK and LL methods suffer from rather severe size distortions, especially when $N$ is large. In view of these findings, we recommend using the proposed method in
Table 2: Coverage Rates of Uniform Confidence Bands ($\rho = 0.5$)

<table>
<thead>
<tr>
<th></th>
<th>Robust T=200</th>
<th>Robust T=500</th>
<th>Robust T=2000</th>
<th>BCCK T=200</th>
<th>BCCK T=500</th>
<th>BCCK T=2000</th>
<th>LL T=200</th>
<th>LL T=500</th>
<th>LL T=2000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case I: N = 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m = 6</td>
<td>0.893</td>
<td>0.916</td>
<td>0.935</td>
<td>0.774</td>
<td>0.774</td>
<td>0.780</td>
<td>0.813</td>
<td>0.814</td>
<td>0.826</td>
</tr>
<tr>
<td>m = 8</td>
<td>0.887</td>
<td>0.914</td>
<td>0.933</td>
<td>0.796</td>
<td>0.807</td>
<td>0.810</td>
<td>0.832</td>
<td>0.839</td>
<td>0.847</td>
</tr>
<tr>
<td>m = 10</td>
<td>0.875</td>
<td>0.908</td>
<td>0.933</td>
<td>0.807</td>
<td>0.820</td>
<td>0.828</td>
<td>0.842</td>
<td>0.847</td>
<td>0.856</td>
</tr>
<tr>
<td>m = 12</td>
<td>0.860</td>
<td>0.906</td>
<td>0.930</td>
<td>0.804</td>
<td>0.831</td>
<td>0.837</td>
<td>0.843</td>
<td>0.861</td>
<td>0.860</td>
</tr>
</tbody>
</table>

Case II: N = 25

| m = 6 | 0.872       | 0.901       | 0.917        | 0.397      | 0.389      | 0.370        | 0.442   | 0.435   | 0.412     |
| m = 8 | 0.875       | 0.907       | 0.929        | 0.454      | 0.447      | 0.456        | 0.504   | 0.489   | 0.493     |
| m = 10| 0.869       | 0.907       | 0.929        | 0.481      | 0.495      | 0.494        | 0.535   | 0.535   | 0.531     |
| m = 12| 0.868       | 0.907       | 0.930        | 0.524      | 0.531      | 0.526        | 0.582   | 0.570   | 0.559     |

Case III: N = 100

| m = 6 | 0.869       | 0.896       | 0.918        | 0.100      | 0.097      | 0.067        | 0.121   | 0.116   | 0.084     |
| m = 8 | 0.867       | 0.902       | 0.922        | 0.136      | 0.136      | 0.129        | 0.158   | 0.155   | 0.146     |
| m = 10| 0.862       | 0.894       | 0.920        | 0.154      | 0.148      | 0.141        | 0.181   | 0.168   | 0.157     |
| m = 12| 0.866       | 0.899       | 0.923        | 0.180      | 0.166      | 0.175        | 0.215   | 0.187   | 0.193     |

Note: This table reports the coverage rates of 95% uniform confidence bands for the proposed (spatio-temporal) robust method, as well as the methods of Belloni, Chernozhukov, Chetverikov, and Kato (2015) and Li and Liao (2020), labeled as robust, BCCK, and LL, respectively. The autoregressive coefficient is set to $\rho = 0.5$. The HAC estimation is performed using the Newey–West estimator with bandwidth $\lfloor 0.75T^{1/3}\rfloor$. The coverage rates are computed based on 10,000 Monte Carlo replications.
panel-data applications. This message extends the lesson of Driscoll and Kraay (1998) to the present new context concerning uniform nonparametric inference.

4 An empirical application on information rigidity

We illustrate the proposed uniform nonparametric inference method in an empirical application concerning the information rigidity of professional forecasters. A large literature in macroeconomics has shown that information frictions can help explain puzzling empirical facts and offer new policy implications different from conventional rational-expectation models; see Mankiw and Reis (2002), Sims (2003), Woodford (2003), Reis (2006), and Mackowiak and Wiederholt (2009), among others. In an influential paper, Coibion and Gorodnichenko (2015) show that both the sticky-information model of Mankiw and Reis (2002) and the noisy-information model of Woodford (2003) and Sims (2003) imply a positive relation between the average ex post forecast errors across agents and their average ex ante forecast revision. These authors also find strong empirical support for this type of information rigidity in a variety of consumer and professional surveys. Set against this background, we revisit Coibion and Gorodnichenko’s (2015) analysis by employing the proposed functional inference method, which allows us to study a more general nonparametric notion of information rigidity. As detailed below, our empirical results provide further support for Coibion and Gorodnichenko’s (2015) findings. In addition, we also conduct a formal nonparametric analysis for state-dependent information rigidity and document pronounced information rigidity conditional on high risk/uncertainty states.

Turning to the details, we first recall the empirical framework developed by Coibion and Gorodnichenko (2015). Let $Z_{k,t+h}$ denote the $k$th economic variable to be forecast, which is realized at time $t + h$. The average time-$t$ forecast across a group of forecasters for this variable is denoted by $F_tZ_{k,t+h}$. The average ex post forecast error
is thus $Z_{k,t+h} - F_t Z_{k,t+h}$, and the average ex ante forecast revision from $t-1$ to $t$ is $F_t Z_{k,t+h} - F_{t-1} Z_{k,t+h}$. Coibion and Gorodnichenko’s (2015) baseline specification is the following linear regression:

$$Z_{k,t+h} - F_t Z_{k,t+h} = c + \beta (F_t Z_{k,t+h} - F_{t-1} Z_{k,t+h}) + \epsilon_{k,h,t},$$

(4.1)

where the error term $\epsilon_{k,h,t}$ is mean-independent of the time-$t$ information, and $\beta$ measures the degree of information rigidity. If the average forecast is rational, one would have $\beta = 0$; otherwise, both sticky-information and noisy-information models imply $\beta > 0$. This testable implication is quite intuitive: Information frictions tend to make the average forecast revision “too conservative” relative to the rational-expectation benchmark. For ease of notation in our discussion below, we shall use $i = (k,h)$ to index both the variable of interest and the forecast horizon, and set $Y_{it} = Z_{k,t+h} - F_t Z_{k,t+h}$, $X_{it} = F_t Z_{k,t+h} - F_{t-1} Z_{k,t+h}$, and $\epsilon_{it} = \epsilon_{k,h,t}$. The specification in (4.1) may be rewritten more concisely as a linear panel regression

$$Y_{it} = c + \beta X_{it} + \epsilon_{it}.$$  

(4.2)

A natural generalization of (4.2) is the nonparametric regression in (2.1), namely, $Y_{it} = g (X_{it}) + \epsilon_{it}$. Here, the counterpart of the slope coefficient $\beta$ is the derivative of the conditional expectation function $g (\cdot)$, denoted $\partial g (\cdot)$. In the nonparametric context, information rigidity implies that $g (\cdot)$ is an increasing function. Compared to the baseline linear specification, the nonparametric model allows the marginal response $\partial g (\cdot)$ to depend on the level of forecast revision itself. This may be formalized in the sticky-information (resp. noisy-information) model if the agents’s updating frequency (resp. Kalman gain) is a function of the revision itself.

We carry out the empirical analysis using the data from the Survey of Professional Forecasters (SPF), which is also the main focus of the Coibion–Gorodnichenko study. For
ease of comparison, we employ exactly the same data as that prior work, obtained from *American Economic Review*’s website. The dataset consists of quarterly time series of forecast errors and revisions from 1969 to 2014 for 5 macroeconomic variables (i.e., GDP price deflator, real GDP, industrial production, housing starts, and the unemployment rate) and 4 horizons (i.e., $h = 0, 1, 2, 3$). The resulting panel corresponds to $N = 20$ and $T = 173$.\(^3\) In this setting, it is clearly implausible to rule out cross-sectional dependence between forecast errors across different variables and/or horizons, or to impose ad hoc “weak” spatial dependence. Since forecasts over multiple horizons are involved, one cannot rule out serial dependence, either (Hansen and Hodrick (1980)). Our proposed method is designed to accommodate this type of general dependence structure for nonparametric analysis, just like how Driscoll and Kraay (1998) handle the classical GMM.

The nonparametric inference is implemented as follows. Since the time series of forecast errors and forecast revisions have quite different scales across variables and horizons, we first make them more comparable by normalizing each individual time series separately so that it is averaged at zero with unit standard deviation. This normalization prevents the pooled nonparametric estimate from being dominated by a few series with large scales. We then follow the same steps as in the simulation. Specifically, to form the series basis, we first transform the covariate $X_{it}$ to $\tilde{X}_{it} = 2\Phi(X_{it}) - 1$, which is strictly increasing in $X_{it}$ and takes value on $[-1, 1]$, and then perform the nonparametric series regression by regressing $Y_{it}$ on a fifth-order Legendre polynomial of the transformed forecast revision $\tilde{X}_{it}$. Our empirical findings are robust with respect to the choice of the number of series.

\(^3\)The SPF dataset also includes shorter time series for additional eight variables starting from 1981. We do not consider those variables because Coibion and Gorodnichenko (2015) document strong heterogeneity in $\beta$ for the totality of 13 variables, which makes the pooled regression more difficult to interpret. On the other hand, for the 5 variables and 4 forecast horizons considered here, Coibion and Gorodnichenko (2015) find only mild evidence for heterogeneity, which can be mainly attributed to the housing starts variable (see their Figure 1). While we focus on the empirical analysis for all 5 variables (including housing starts) in the main text, we have also performed a similar analysis with housing starts excluded. This does not change our empirical message and so the details for that robustness check are relegated to the Supplemental Appendix.
Figure 1: Nonparametric Estimation of Information Rigidity

Note: The left panel plots the nonparametric estimate of the conditional mean function of the forecast error given the transformed forecast revision, and the right panel plots the nonparametric estimate of its derivative. Each time series of forecast error or forecast revision is normalized to have zero sample mean and unit standard deviation. The transformed forecast revision is defined by transforming the normalized data via $x \mapsto 2\Phi(x) - 1$ so that it takes values on the $[-1, 1]$ interval. The nonparametric series estimation is implemented using Legendre polynomials up to the fifth order. The 90% uniform confidence band is computed using Algorithm 1 (resp. Algorithm 2) for the left (resp. right) panel, where the Newey–West bandwidth is set to $\lfloor 0.75T^{1/3} \rfloor$.

We now turn to the empirical results. On the left panel of Figure 1, we plot the nonparametric estimate of the conditional mean function of the forecast error $Y_{it}$ given the transformed forecast revision $\tilde{X}_{it}$, along with its 90% uniform confidence band computed according to Algorithm 1. We first observe that the functional estimate is statistically significantly different from zero, which provides nonparametric evidence against the hypothesis that the SPF forecasts are fully rational. Importantly, the estimated conditional expectation function appears to be an increasing function in the forecast revision, which,
as mentioned above, may be interpreted as nonparametric evidence for the presence of information rigidity. To see this more clearly, on the right panel of Figure 1, we plot the nonparametric estimate of the derivative function, together with its 90% uniform confidence band computed according to Algorithm 2. The plot reveals that the derivative estimate is indeed almost always positive, and the nonparametric estimate as a whole is significantly different from zero. Overall, these findings provide strong support for those of Coibion and Gorodnichenko (2015): The information rigidity documented in the prior work is not solely driven by the linear specification but holds quite robustly in a nonparametric setting; moreover, information rigidity is not only an “on-average” phenomenon (as summarized by the scalar $\beta$), but also appears to hold “uniformly” in a functional sense across different levels of forecast revision.

We may also formally test whether the linear specification (4.2) is in fact compatible with observed data. This type of specification test can be easily carried out using the proposed method. To do so, we first estimate the linear model and obtain the residual; we then nonparametrically regress the residuals on the covariate. If the linear model is correctly specified, then the nonparametrically estimated conditional mean function of the linear-regression residual should be statistically zero; otherwise, the linear specification should be rejected. Figure 2 plots the estimated conditional mean function of the residual and the associated 90% uniform confidence band. Since the confidence band always covers zero, we cannot reject the linear specification in (4.2). This suggests that the rigidity parameter $\beta$ is likely constant across different levels of forecast revision and so Coibion and Gorodnichenko’s (2015) linear specification (with constant $\beta$) is indeed adequate from this perspective.

That being said, the level of information rigidity may vary with the other types of economic state variables. Coibion and Gorodnichenko (2015) provide secular evidence that information rigidity rose systematically over the course of the Great Moderation, which
Figure 2: Test for Linear Specification

Note: The figure plots the nonparametric estimate of the conditional mean function of the linear-regression residual given the transformed forecast revision. Each time series of forecast error and forecast revision is normalized to have zero sample mean and unit standard deviation. The residuals are obtained from a pooled linear panel regression according to (4.2). The transformed forecast revision is defined by transforming the normalized data via $x \mapsto 2\Phi(x) - 1$ so that it takes values on the $[-1, 1]$ interval. The nonparametric series estimation is implemented using Legendre polynomials up to the fifth order. The 90% uniform confidence band is computed using Algorithm 1, where the Newey–West bandwidth is set to $\lfloor 0.75T^{1/3} \rfloor$.

may be explained by the less attention paid by economic agents during tranquil times. However, this may not be the full story for at least two reasons. Firstly, while macroeconomic volatility is low during the Great Moderation period, the U.S. economy also experienced several high risk/uncertainty episodes including the 1987 Black Monday stock market crash, the Gulf War, the Dot-com bubble, the September 11 attacks, the Second Gulf War, and several presidential elections. Secondly, although economic agents may allocate more attention to aggregate state variables during more volatile or uncertain periods, the high risk/uncertainty also makes the learning about the underlying economic
states more difficult. Indeed, a Bayesian learner would assign less weight on a noisier signal; moreover, an ambiguity-averse agent may be “paralyzed by fear” in the presence of high uncertainty (Ilut, Valchev, and Vincent (2020)). Economic agents may thus underreact to new information even if they allocate more attention to it, which is in line with Tversky and Kahneman’s (1974) notion of conservatism in their study of judgment under uncertainty. By this logic, the level of information rigidity could be high during high risk/uncertainty periods. Whether this is the case warrants some empirical exploration.

To formally investigate this question, we generalize the linear specification (4.2) by modeling $\beta$ as a state-dependent function. Specifically, we consider

$$Y_{it} = c + \beta(U_t)X_{it} + \epsilon_{it},$$

(4.3)

where $U_t$ is a measure for market-level risk or uncertainty. A key advantage of the nonparametric series regression is that it can be easily adapted to estimate this type of “partially linear” model (in a multiplicative form). The idea is to approximate $\beta(\cdot)$ with a linear combination $P(\cdot)^\top \gamma$ of approximating functions, and estimate the model by regressing $Y_{it}$ on the constant term and the regressor vector $P(U_t)X_{it}$. With $\hat{\gamma}$ denoting the regression coefficient for $P(U_t)X_{it}$, the series estimator for $\beta(\cdot)$ is given by $P(\cdot)^\top \hat{\gamma}$. Algorithm 1 can be used to construct the uniform confidence band for $\beta(\cdot)$ with one modification, that is, to replace $\hat{A}$ with the HAC estimator for $\hat{\gamma}$.

Figure 3 plots the nonparametric estimate of the $\beta(\cdot)$ function and its 90% uniform confidence band when $U_t$ is the log volatility of the U.S. stock market portfolio computed as the logarithm of the standard deviation of daily returns (in percentage) over the preceding month.\footnote{The daily returns of the market portfolio is obtained from Kenneth French’s website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html} From the figure, we see that the estimate of information rigidity takes the highest and the most statistically significant values when the stock market volatility is
Figure 3: Information Rigidity and Stock Market Volatility

Note: This figure plots the nonparametric estimate of the functional coefficient $\beta(\cdot)$ in the specification $Y_{it} = c + \beta(U_t) X_{it} + \epsilon_{it}$, where $U_t$ is the log volatility of the stock market portfolio, computed as the standard deviation of daily portfolio returns (in percentage) over the preceding month. The nonparametric series estimation is implemented by approximating $\beta(\cdot)$ with a fifth-order Legendre polynomial. The 90% uniform confidence band is computed using Algorithm 1, where the Newey–West bandwidth is set to $\left\lfloor 0.75T^{1/3} \right\rfloor$.

high, but is largely insignificant over the low-volatility region. This finding suggests that forecasters’s under-reaction during volatile periods (in the stock market) may be a key factor for explaining the overall information rigidity. This result is also in line with Coibion and Gorodnichenko’s (2015) finding that the degree of information rigidity is high when the economy enters a recession.

To further corroborate this finding, we repeat the above analysis but instead taking $U_t$ as either the macro uncertainty or the financial uncertainty measure constructed by Jurado, Ludvigson, and Ng (2015). The correlation coefficient between these two uncertainty measures is 0.51, and their correlations with the log volatility of the market portfolio are 0.31 and 0.69, respectively. Not surprisingly, the financial uncertainty measure is highly corre-
Figure 4: Information Rigidity and Uncertainty

Note: The left (resp. right) panel plots the nonparametric estimate of the functional coefficient $\beta(\cdot)$ in the specification $Y_{it} = c + \beta(U_t)X_{it} + \epsilon_{it}$, where $U_t$ is the macro (resp. financial) uncertainty measure constructed by Jurado, Ludvigson, and Ng (2015). The nonparametric series estimation is implemented by approximating $\beta(\cdot)$ with a fifth-order Legendre polynomial. The 90% uniform confidence band is computed using Algorithm 1, where the Newey–West bandwidth is set to $\left\lfloor 0.75T^{1/3} \right\rfloor$.

lated with market volatility; but their dependence is far from perfect because the former covers a much broader range of financial information. Figure 4 depicts how information rigidity depends on these uncertainty measures. The nonparametric estimate shown on the left panel does not reveal a clear relationship between information rigidity and macro uncertainty. Nevertheless, the estimated $\beta(\cdot)$ is mostly positive, and the functional estimator is statistically different from zero. From the right panel, we see that the degree of information rigidity tends to increase with financial uncertainty, and the overall pattern is quite similar to Figure 3. This further confirms the interesting link between information rigidity and the level of risk/uncertainty in the financial market. This is perhaps not surprising, because the financial market is arguably the most important mechanism for aggregating and
disseminating information. Asset prices contain important forward-looking information concerning various macroeconomic variables such as interest rates and inflation, and are closely watched by economic agents. Therefore, when the financial market experiences high volatility or uncertainty, a forecaster may find it more difficult to separate fundamental information from the “noise” in asset prices (Grossman and Stiglitz (1980), Kyle (1985)). This type of friction may help explain the high level of information rigidity seen in our nonparametric estimates.

In summary, the empirical illustration above demonstrates how the proposed uniform inference method may be used in applied work. We see that this tool is quite versatile: In addition to conducting inference for the conditional mean function itself, one may also apply the method to study the derivative function, perform specification tests (e.g., testing for linear specification), and estimate state-dependent response. Our empirical findings provide nonparametric evidence for the presence of information rigidity, which further supports the main findings of Coibion and Gorodnichenko (2015). We also provide new empirical evidence for the positive relationship between information rigidity and financial risk/uncertainty.

5 Conclusion

Nonparametric regressions offer flexible empirical designs but need more data for informative inference. This need could hinder macroeconomic applications in which the number of observations for a typical time series is often in the low hundreds. A reasonable and oft-used empirical strategy to overcome this issue is to pool the richer information from a panel. The related inference should be done carefully due to the presence of serial and cross-sectional dependence in the data. The proposed uniform nonparametric inference method readily accommodates general spatio-temporal dependence. It may be used to
make functional inference concerning the conditional mean function in panel-data settings with a large $T$, regardless whether $N$ is small or large.

APPENDIX

PROOF OF THEOREM 1. By Assumption 1(i), $(T_n^{-1/2}H_{n,t})_{t \geq 1}$ forms a martingale difference array with respect to the filtration $\{\mathcal{F}_t\}_{t \geq 1}$. Moreover, by Assumptions 1(ii, iv), Assumption 1 in Li and Liao (2020) holds for $T_n^{-1/2}H_{n,t}$. Therefore, applying Theorem 1 of Li and Liao (2020), we have

$$\|S_n - \xi_n\| = O_p(m_n^{1/2}r_n^{1/2} + (B_n m_n)^{1/3}) \quad (A.1)$$

where $B_n \equiv \sum_{t=1}^{T_n} \mathbb{E}[\|T_n^{-1/2}H_{n,t}\|^3]$ and $\xi_n$ is an $m_n$-dimensional random vector with distribution $\mathcal{N}(0, T_n^{-1} \sum_{t=1}^{T_n} \text{Var}[H_{n,t}])$. Next, we can bound $B_n$ as follows:

$$\sum_{t=1}^{T_n} \mathbb{E}[\|T_n^{-1/2}H_{n,t}\|^3] \leq \sum_{t=1}^{T_n} \mathbb{E} \left[ \left( \sum_{j=1}^{m_n} (T_n^{-1/2}H_{n,t}^{(j)})^2 \right)^{3/2} \right] \leq T_n^{-3/2} m_n^{1/2} \sum_{t=1}^{T_n} \sum_{j=1}^{m_n} \mathbb{E}[|H_{n,t}^{(j)}|^3] = O(T_n^{-1/2} m_n^{3/2}),$$

where the inequality is by Hölder’s inequality and the last equality is by Assumption 1(iii). Plugging this estimate into (A.1), we readily deduce the assertion of Theorem 1. Q.E.D.

PROOF OF THEOREM 2. Under condition (i), $(T_n^{-1/2}H_{n,t})_{t \geq 1}$ is a mixingale satisfying Assumption 5 of Li and Liao (2020). Since $c_n = O(1)$ and the other sufficient conditions
of Theorem 4 of Li and Liao (2020) are maintained in conditions (ii, iii), we can apply this theorem and Theorem 1 to show that

\[ \| S_n - \xi_n \| = O_p \left( m_n^{1/2} T_n^{-1/2} \right) + O_p \left( m_n^{1/2} r_n^{1/2} + m_n^{5/6} T_n^{-1/6} \right) + O_p \left( m_n T_n^{-1/2} + m_n^{3/2} T_n^{-1} \right) \]

(A.2)

where \( \xi_n \) is an \( m_n \)-dimensional random vector with distribution \( \mathcal{N}(0, \text{Var}[S_n]) \). Since \( m_n = o(T_n) \) by condition (iv), the assertion of Theorem 2 follows from (A.2). \( Q.E.D. \)

References


CHEN, X., AND T. M. CHRISTENSEN (2015): “Optimal Uniform Convergence Rates and Asymp-


