Uniform Nonparametric Inference for Spatially Dependent Panel Data

Jia Li* Zhipeng Liao† Wenyu Zhou‡

December 23, 2021

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 in two empirical settings. One concerns the nonparametric relationship between asset price volatility and trading volume as depicted by the mixture of distribution hypothesis. The other pertains to testing the rationality of survey-based forecasts, in which we document nonparametric evidence for information rigidity among professional forecasters, offering new support for sticky-information and noisy-information models in macroeconomics.

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

*School of Economics, Singapore Management University, Singapore; e-mail: jiali@smu.edu.sg.

†Department of Economics, UCLA, Log 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 et al. (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 empirical settings (Bertrand et al. (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 (Eubank and Spiegelman (1990), Andrews (1991a), Newey (1997), Huang (1998, 2003)), 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, Chernozhukov et al. (2013) and Belloni et al. (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. Li and Liao (2020) extend this theory to a general time-series setting for heterogeneous mixingales, which permits a broader range of empirical 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, for which it is important to accommodate different types of spatio-temporal dependence encountered by empirical researchers (Bertrand et al. (2004), Petersen (2009)). In the baseline context of linear panel regressions, this mainly manifests as alternative ways of computing standard errors. One 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, which are our main empirical focus. 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 statistical 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 extremum estimator, but we focus on how to make uniform functional inference for the conditional expectation function.

To illustrate the usefulness of the proposed procedure, we conduct two empirical applications. The first concerns the functional relationship between price volatility and trading volume of financial assets (Clark (1973), Tauchen and Pitts (1983), Andersen (1996)). Specifically, we study how the volume-volatility relationship estimated using a recent panel consisting of all stocks listed in the U.S. equity market has changed after the breakout of the COVID-19 pandemic, and document a significant higher market impact of trades during the post-COVID period. In the second application, we study the rationality of survery-based forecasts, by estimating the nonparametric relationship between the average ex post forecast error and ex ante forecast revision in the Survey of Professional Forecasters (SPF). Consistent with the theoretical prediction from the information-rigidity theory (Mankiw and Reis (2002), Sims (2003), Woodford (2003), Reis (2006), Coibion and Gorodnichenko (2015)), our nonparametric estimate of the conditional mean function of the forecast error is increasing in the forecast revision, and so, provides robust nonparametric evidence for the presence of sticky or noisy information.

The remainder of the paper is organized as follows. Section 2 presents our uniform nonparametric inference method for panel data. The empirical applications are provided in Section 3. Section 4 concludes. All proofs are in the Appendix. The online supplemental appendix reports the finite-sample performance of the proposed method in a Monte Carlo study.

2 The statistical method

In this section, we present the uniform nonparametric inference procedure. We describe the setting and some relevant background in Section 2.1. Section 2.2 presents new 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 $\left(Y_{it}, \boldsymbol{X}_{it}^{\top}\right)_{1\leq i\leq N, 1\leq t\leq T}$ where Y_{it} is a scalar-valued dependent variable and the covariate $\boldsymbol{X}_{it}$ takes value in a compact set $\boldsymbol{\mathcal{X}} \subseteq \mathbb{R}^{d}$. 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 $\boldsymbol{X}_{it}$, denoted by $g(\boldsymbol{x}) \equiv \mathbb{E}[Y_{it}|\boldsymbol{X}_{it} = \boldsymbol{x}]$ for $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$. Setting the disturbance term as $\epsilon_{it} \equiv Y_{it} - g(\boldsymbol{X}_{it})$, we may equivalently state the problem as a nonparametric regression

$$Y_{it} = g(\boldsymbol{X}_{it}) + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it}|\boldsymbol{X}_{it}] = 0. \tag{2.1}$$

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

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

Our procedure is built on the series regression method (Eubank and Spiegelman (1990), Andrews (1991a), Newey (1997), Huang (1998, 2003)). Under the series approach, the nonparametric estimation can be performed by running a (pooled) least-squares regression of Y_{it} on a collection of approximating functions of $\boldsymbol{X}_{it}$. Specifically, consider a column vector of approximating functions $\boldsymbol{P}(\cdot) = (p_j(\cdot))_{1 \leq j \leq m}$, which may be polynomials, splines, trigonometric functions, wavelets, etc.; see Chen (2007) for a comprehensive review. Regressing Y_{it} on $\boldsymbol{P}(\boldsymbol{X}_{it})$ yields the regression coefficient

$$\widehat{\boldsymbol{b}} \equiv \left(\sum_{t=1}^{T} \sum_{i=1}^{N} \boldsymbol{P}(\boldsymbol{X}_{it}) \boldsymbol{P}(\boldsymbol{X}_{it})^{\top} \right)^{-1} \left(\sum_{t=1}^{T} \sum_{i=1}^{N} \boldsymbol{P}(\boldsymbol{X}_{it}) Y_{it} \right), \tag{2.3}$$

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

$$\widehat{g}(\cdot) \equiv \boldsymbol{P}(\cdot)^{\top} \widehat{\boldsymbol{b}}. \tag{2.4}$$

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 $m \to \infty$ asymptotically, 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 $g(\boldsymbol{x})$ at a given point $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ 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., $\widehat{g}(\cdot)$ 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 $\widehat{\boldsymbol{b}}$, because its dimensionality is divergent asymptotically. This in turn leads to difficulties for establishing the asymptotic Gaussianity of the functional estimator $\widehat{g}(\cdot)$.

In a cross-sectional setting (i.e., $T = 1$) with independent data, Belloni et al. (2015) show that the aforementioned non-Donsker issue may be addressed by using a strong Gaussian approximation theory. For ease of discussion, denote $\boldsymbol{h}_{it} \equiv \boldsymbol{P}(\boldsymbol{X}_{it})\, \epsilon_{it}$, so that the score vector for the (single cross-section) series regression may be written as $N^{-1/2} \sum_{i=1}^{N} \boldsymbol{h}_{i1}$. 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 $\boldsymbol{\xi}_N$ with the same variance-covariance matrix, that is,

$$\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{h}_{i1} - \boldsymbol{\xi}_N \right\| = o_p\left(1\right), \ \boldsymbol{\xi}_N \sim \mathcal{N}\left(0, \frac{1}{N} \sum_{i=1}^{N} \mathrm{Var}\left[\boldsymbol{h}_{i1}\right]\right), \tag{2.5}$$

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

$$\left\| N^{1/2}\left(\widehat{\boldsymbol{b}} - \boldsymbol{b}^*\right) - \boldsymbol{Q}^{-1}\boldsymbol{\xi}_N \right\| = o_p(1),$$

where $\boldsymbol{b}^*$ is the population regression coefficient and $\boldsymbol{Q}$ is the population Gram matrix $\mathbb{E}[N^{-1}\sum_{i=1}^{N} \boldsymbol{P}\left(\boldsymbol{X}_{i1}\right) \boldsymbol{P}\left(\boldsymbol{X}_{i1}\right)^{\top}]$. The coupling of the regression coefficient in turn implies that the scaled estimation error function $N^{1/2}(\widehat{g}\left(\cdot\right) - g\left(\cdot\right))$ may be coupled by a Gaussian process $\boldsymbol{P}\left(\cdot\right)^{\top} \boldsymbol{Q}^{-1}\boldsymbol{\xi}_N$, which can then be used to construct uniform confidence bands for $g\left(\cdot\right)$.

The logic above reveals that the key to establish a uniform inference theory is the growing-dimensional Gaussian coupling for the score vector as described in (2.5). Along this line, 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 ith time series admits a Gaussian coupling in the form of

$$\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{h}_{it} - \boldsymbol{\xi}_T^{(i)} \right\| = o_p\left(1\right), \ \boldsymbol{\xi}_T^{(i)} \sim \mathcal{N}\left(0, \mathrm{Var}\left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{h}_{it}\right]\right). \tag{2.6}$$

Not surprisingly, the variance-covariance matrix of the coupling variable $\boldsymbol{\xi}_T^{(i)}$ is generally the long-run variance-covariance matrix of the $(\boldsymbol{h}_{it})_{t\geq 1}$ 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 the aforementioned theory to the panel-data setting, 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 now well recognized to be 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

$$\frac{1}{N\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{N} \boldsymbol{h}_{it} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{H}_t, \quad \text{where} \quad \boldsymbol{H}_t \equiv \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{h}_{it}. \tag{2.7}$$

This simple rewriting highlights the fact that the analysis for spatially dependent panels closely resembles the (seemingly) simpler time-series problem, except that the $\boldsymbol{H}_t$ time series is now "generated" as a cross-sectional average of the unit-specific influence function $\boldsymbol{h}_{it}$. 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 $\boldsymbol{h}_{it}$ 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 $\mathcal{F}_t$. We do not always assume that $\boldsymbol{h}_{it}$ is measurable with respect to $\mathcal{F}_t$, but the filtration is useful to specify the serial dependence of $\boldsymbol{h}_{it}$.

Our goal is to construct a Gaussian coupling for a sequence of m_n-dimensional random vectors given by $\boldsymbol{S}_n \equiv T_n^{-1/2} \sum_{t=1}^{T_n} a_N^{-1} \sum_{i=1}^{N} \boldsymbol{h}_{it}$. We stress that the key novelty of this type of results is that it concerns the asymptotic normality of *growing-dimensional* statistics, which is very different from conventional central limit theorems for fixed-dimensional statistics or tight empirical processes. Like Driscoll and Kraay (1998), we rewrite $\boldsymbol{S}_n$ as

$$\boldsymbol{S}_n \equiv \frac{1}{\sqrt{T_n}} \sum_{t=1}^{T_n} \boldsymbol{H}_{n,t}, \quad \text{where} \quad \boldsymbol{H}_{n,t} \equiv \frac{1}{a_N} \sum_{i=1}^{N} \boldsymbol{h}_{it}.$$

The n subscript in $\boldsymbol{H}_{n,t}$ highlights the fact that it generally forms a triangular array. The normalizing sequence a_N is introduced to ensure that $a_N^{-1}\sum_{i=1}^N \boldsymbol{h}_{it}$ is non-degenerate. For example, one may set $a_N = N$ if the $\boldsymbol{h}_{it}$ 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 $\boldsymbol{h}_{it}$ 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 $(\boldsymbol{h}_{it})_{t\geq 1}$ 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 $(\boldsymbol{H}_{n,t})_{t\geq 1}$ 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 $\boldsymbol{H}_{n,t}^{(j)}$ denotes the jth component of $\boldsymbol{H}_{n,t}$.

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

$$\left\| \frac{1}{T_n} \sum_{t=1}^{T_n'} \left(\mathbb{E}[\boldsymbol{H}_{n,t}\boldsymbol{H}_{n,t}^\top|\mathcal{F}_{t-1}] - \mathbb{E}[\boldsymbol{H}_{n,t}\boldsymbol{H}_{n,t}^\top] \right) \right\|_S = O_p(r_n), \tag{2.8}$$

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

Condition (i) of Assumption 1 states that $\boldsymbol{h}_{it}$ forms an MDS for each i with respect to the filtration $\mathcal{F}_t$, which further implies that $\boldsymbol{H}_{n,t}$ is a martingale difference array. Condition (ii) requires that the variance-covariance matrix of $\boldsymbol{H}_{n,t}$ is non-degenerate, and condition (iii) further imposes a bound on its third moment. Condition (iv) mainly requires that the conditional variance-covariance matrix $\mathbb{E}[\boldsymbol{H}_{n,t}\boldsymbol{H}_{n,t}^\top|\mathcal{F}_{t-1}]$ 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 $\boldsymbol{H}_{n,t}$ 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.

Theorem 1, below, establishes the Gaussian coupling for the $\boldsymbol{S}_n$ statistic when $\boldsymbol{h}_{it}$ forms an MDS.

Theorem 1. *Under Assumption 1, there exists a sequence $\boldsymbol{\xi}_n$ of m_n-dimensional random vectors with distribution $\mathcal{N}(0, T_n^{-1}\sum_{t=1}^{T_n}\mathrm{Var}[\boldsymbol{H}_{n,t}])$ such that*

$$\|\boldsymbol{S}_n - \boldsymbol{\xi}_n\| = O_p(m_n^{1/2}r_n^{1/2} + T_n^{-1/6}m_n^{5/6}). \tag{2.9}$$

A couple of remarks are in order. Firstly, we note that the variance-covariance matrix of the coupling variable is $T_n^{-1}\sum_{t=1}^{T_n}\mathrm{Var}[\boldsymbol{H}_{n,t}]$, which does not involve any autocovariance, because $\boldsymbol{h}_{it}$ 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 Nm_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 $\boldsymbol{H}_{n,t} = a_N^{-1} \sum_{i=1}^{N} \boldsymbol{h}_{it}$ 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 $k \geq 0$,

$$\left\| \mathbb{E}[\boldsymbol{H}_{n,t}^{(j)} | \mathcal{F}_{t-k}] \right\|_q \leq \bar{c}_n \psi_k, \quad \left\| \boldsymbol{H}_{n,t}^{(j)} - \mathbb{E}[\boldsymbol{H}_{n,t}^{(j)} | \mathcal{F}_{t+k}] \right\|_q \leq \bar{c}_n \psi_{k+1}, \tag{2.10}$$

where $\|\cdot\|_q$ denotes the L_q-norm of a random variable for some $q \geq 1$, and the constants $\bar{c}_n$ and ψ_k control the magnitude and the serial dependence of the $(\boldsymbol{H}_{n,t})_{t \geq 1}$ variables, respectively. Note that if $\boldsymbol{H}_{n,t}$ forms a martingale difference array and each of its entries has bounded qth 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 applications; 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 $\boldsymbol{H}_{n,t}$, instead of specifying conditions on the $\boldsymbol{h}_{it}$ variables (though this is often easy to do using routine techniques).

We extend Theorem 1 to the more general mixingale case via a martingale approximation. Specifically, under the maintained assumption $\sum_{k \geq 0} \psi_k < \infty$, it can be shown that

$$\left\| \boldsymbol{S}_n - \frac{1}{\sqrt{T_n}} \sum_{t=1}^{T_n} \boldsymbol{H}_{n,t}^* \right\| = O_p(T_n^{-1/2} m_n^{1/2}), \tag{2.11}$$

where

$$\boldsymbol{H}_{n,t}^* \equiv \sum_{s=-\infty}^{\infty} \left\{ \mathbb{E}\left[\boldsymbol{H}_{n,t+s} | \mathcal{F}_t \right] - \mathbb{E}\left[\boldsymbol{H}_{n,t+s} | \mathcal{F}_{t-1} \right] \right\}. \tag{2.12}$$

Observe that $\boldsymbol{H}_{n,t}^*$ forms a martingale difference array and so $T_n^{-1/2}\sum_{t=1}^{T_n}\boldsymbol{H}_{n,t}^*$ admits a strong Gaussian coupling by Theorem 1. Since $\boldsymbol{S}_n$ can be approximated by $T_n^{-1/2}\sum_{t=1}^{T_n}\boldsymbol{H}_{n,t}^*$ up to a relatively small $O_p(T_n^{-1/2}m_n^{1/2})$ error, this further implies that $\boldsymbol{S}_n$ also admits a Gaussian coupling. Theorem 2, below, formalizes this logic.

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

$$\|\boldsymbol{S}_n - \boldsymbol{\xi}_n\| = O_p(m_n^{1/2}r_n^{1/2} + T_n^{-1/6}m_n^{5/6}). \tag{2.13}$$

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

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 $g\left(\cdot\right)$. With the main goal of facilitating practical applications, we shall focus on discussing the implementation and the underlying intuition. The heuristics may be formalized by adapting routine theoretical arguments as in Belloni et al. (2015) and Li and Liao (2020).

Turning to the details, we first recall from (2.3) that $\widehat{\boldsymbol{b}}$ is the least-squares coefficient obtained by regressing Y_{it} on $\boldsymbol{P}(\boldsymbol{X}_{it})$, and $\widehat{g}(\cdot) = \boldsymbol{P}(\cdot)^{\top}\widehat{\boldsymbol{b}}$ is the nonparametric series estimator for $g(\cdot)$. When the number of series terms m is large, we have $g(\cdot) \approx \boldsymbol{P}(\cdot)^{\top}\boldsymbol{b}^{*}$ for some "population" regression coefficient $\boldsymbol{b}^{*}$ and so $Y_{it} \approx \boldsymbol{P}(\boldsymbol{X}_{it})^{\top}\boldsymbol{b}^{*} + \epsilon_{it}$. Hence, with $\widehat{\boldsymbol{Q}} \equiv T^{-1}\sum_{t=1}^{T} N^{-1}\sum_{i=1}^{N} \boldsymbol{P}(\boldsymbol{X}_{it})\boldsymbol{P}(\boldsymbol{X}_{it})^{\top}$, we have

$$\widehat{\boldsymbol{b}} - \boldsymbol{b}^{*} \approx \widehat{\boldsymbol{Q}}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{P}(\boldsymbol{X}_{it})\,\epsilon_{it}\right), \tag{2.14}$$

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 of $\widehat{\boldsymbol{b}}$ may be established by applying the aforementioned Gaussian coupling theorems for the panel $\boldsymbol{h}_{it} = \boldsymbol{P}(\boldsymbol{X}_{it})\,\epsilon_{it}$. To do so, we set $\boldsymbol{H}_{t} = a_{N}^{-1}\sum_{i=1}^{N}\boldsymbol{h}_{it}$ as in Section 2.2 (the n subscript is omitted here for simplicity). With this notation, we rewrite (2.14) as

$$\frac{T^{1/2}N}{a_{N}}\left(\widehat{\boldsymbol{b}} - \boldsymbol{b}^{*}\right) \approx \widehat{\boldsymbol{Q}}^{-1} \times \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{H}_{t}. \tag{2.15}$$

Theorem 2 above shows that $T^{-1/2}\sum_{t=1}^{T}\boldsymbol{H}_{t}$ may be strongly approximated by a Gaussian vector $\boldsymbol{\xi}_{n} \sim \mathcal{N}(0, \boldsymbol{A})$, where

$$\boldsymbol{A} \equiv \mathrm{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{H}_{t}\right] = \mathrm{Var}\left[\frac{1}{a_{N}\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{N}\boldsymbol{P}(\boldsymbol{X}_{it})\epsilon_{it}\right]. \tag{2.16}$$

The estimation error in $\widehat{\boldsymbol{b}}$ thus admits the following Gaussian approximation in distribution:

$$\frac{T^{1/2}N}{a_{N}}\left(\widehat{\boldsymbol{b}} - \boldsymbol{b}^{*}\right) \overset{d}{\approx} \left(\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}^{-1}\right)^{1/2}\boldsymbol{\mathcal{N}}_{m}^{*}, \tag{2.17}$$

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

$$\frac{T^{1/2}N}{a_N}\left(\widehat{g}(\cdot) - g(\cdot)\right) \overset{d}{\approx} \boldsymbol{P}(\cdot)^{\top}\left(\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}^{-1}\right)^{1/2}\boldsymbol{\mathcal{N}}_{m}^{*}. \tag{2.18}$$

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

$$\sigma(\cdot) \equiv \sqrt{\boldsymbol{P}(\cdot)^{\top}\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{P}(\cdot)}. \tag{2.19}$$

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

$$\begin{cases} \widehat{\boldsymbol{A}}(a_N) \equiv \displaystyle\sum_{s=-(M_n-1)}^{M_n-1}\frac{|M_n - s|}{M_n}\widehat{\boldsymbol{\Gamma}}_s(a_N), \quad \text{where} \\ \widehat{\boldsymbol{\Gamma}}_s(a_N) \equiv \displaystyle\frac{1}{T}\sum_{t=\max\{1,1-s\}}^{\min\{T-s,T\}}\left(\frac{1}{a_N}\sum_{i=1}^{N}\boldsymbol{P}(\boldsymbol{X}_{it})\widehat{\epsilon}_{it}\right)\left(\frac{1}{a_N}\sum_{i=1}^{N}\boldsymbol{P}(\boldsymbol{X}_{it+s})\widehat{\epsilon}_{it+s}\right)^{\top}, \end{cases} \tag{2.20}$$

$\widehat{\epsilon}_{it} = Y_{it} - \widehat{g}(\boldsymbol{X}_{it})$, 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 $\widehat{\boldsymbol{A}}(a_N)$. However, since $m \to \infty$ in the present nonparametric setting, we may 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 $\boldsymbol{P}(\boldsymbol{X}_{it})\epsilon_{it}$ forms an MDS, then the user does not need to include sample autocovariances in $\widehat{\boldsymbol{A}}(a_N)$, which amounts to setting $M_n = 1$. Equipped with $\widehat{\boldsymbol{A}}(a_N)$, we estimate the standard error function $\sigma(\cdot)$ defined in (2.19) via its sample analogue

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

For notational simplicity in the discussion below, we omit a_N in the notation when $a_N = N$; in particular, we write $\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{A}}(N)$ and $\widehat{\sigma}(\cdot) = \widehat{\sigma}(\cdot;N)$. It is worth noting that $\widehat{\boldsymbol{A}}$ and $\widehat{\sigma}(\cdot)$ 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

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

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 $\widehat{\sigma}(\cdot;a_N) = Na_N^{-1}\widehat{\sigma}(\cdot)$ by definition, and hence, we may rewrite $\hat{\tau}$ as $\hat{\tau} \equiv \sup_{\boldsymbol{x}\in\boldsymbol{\mathcal{X}}} T^{1/2}\left|\widehat{g}(\boldsymbol{x}) - g(\boldsymbol{x})\right|/\widehat{\sigma}(\boldsymbol{x})$. In view of the Gaussian approximation in (2.18), we may approximate the distribution of $\hat{\tau}$ via that of

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

for which a feasible approximation may be further constructed as

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

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 $g(\cdot)$.

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 $\boldsymbol{P}(\boldsymbol{X}_{it})$ and obtain $\widehat{\boldsymbol{b}}$ as described in (2.3). Set the nonparametric estimator $\widehat{g}(\cdot) = \boldsymbol{P}(\cdot)^\top \widehat{\boldsymbol{b}}$.

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

Step 3. Draw $\boldsymbol{\mathcal{N}}_m^*$ from the m-dimensional standard normal distribution many times. For each draw, compute

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

where the supremum may be computed on a discretized mesh of $\boldsymbol{\mathcal{X}}$. 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 $[\widehat{g}(\cdot) - cv_{1-\alpha} T^{-1/2} \widehat{\sigma}(\cdot), \widehat{g}(\cdot) + cv_{1-\alpha} T^{-1/2} \widehat{\sigma}(\cdot)]$. $\square$

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 $\widehat{\boldsymbol{b}}$, $\widehat{\boldsymbol{Q}}$, and $\widehat{\boldsymbol{A}}$ as described in Algorithm 1. Set

$$\partial \widehat{g}(\cdot) = \partial \boldsymbol{P}(\cdot)^\top \widehat{\boldsymbol{b}}, \quad \tilde{\sigma}(\cdot) = \sqrt{\partial \boldsymbol{P}(\boldsymbol{x})^\top \left(\widehat{\boldsymbol{Q}}^{-1} \widehat{\boldsymbol{A}} \widehat{\boldsymbol{Q}}^{-1} \right) \partial \boldsymbol{P}(\boldsymbol{x})}.$$

Step 2. Draw $\boldsymbol{\mathcal{N}}_m^*$ from the m-dimensional standard normal distribution many times. For each draw, compute

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

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

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

The statistical procedures described in Algorithms 1 and 2 can be readily implemented using an accompanying Stata package available at the authors' websites. We also note that, since the underlying asymptotic theory is developed in a general setting with triangular arrays, the proposed method can be easily extended (mainly at the cost of more complicated notation) to the setting in which the number of observations in each cross-section, say N_t, depends on t. The only modification needed is to replace the cross-sectional average $N^{-1}\sum_{i=1}^{N}$ with $N_t^{-1}\sum_{i=1}^{N_t}$ when computing various sample-average statistics. This more general setting is accommodated in the aforementioned Stata package as well.

3 Empirical applications

In this section, we demonstrate how the proposed uniform inference method may be used in empirical work via two examples: The first concerns the nonparametric relationship between the volatility and trading volume of financial assets, and the second pertains to testing the rationality of survey-based forecasts.

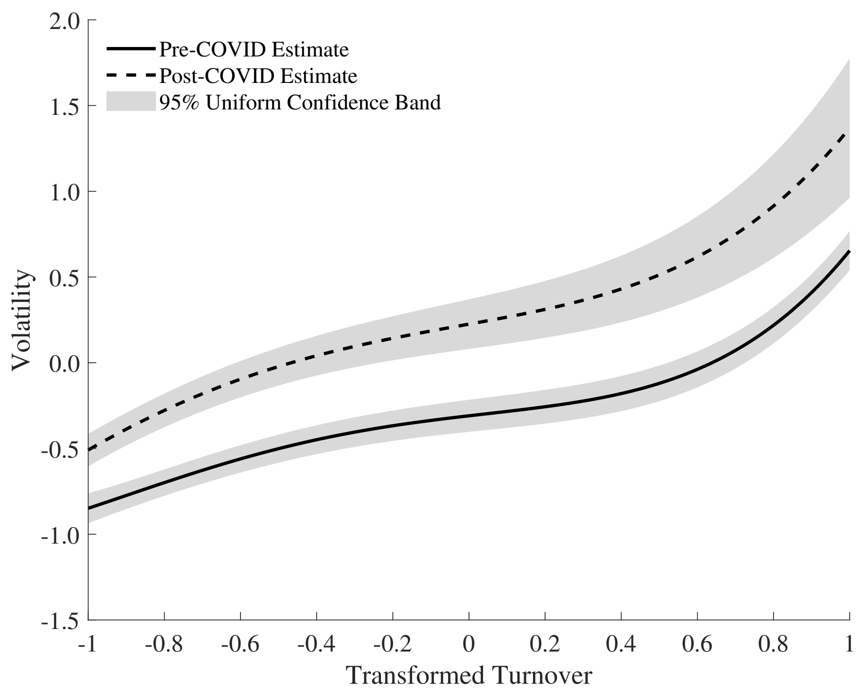
3.1 Volume-volatility relationship during COVID-19

A large literature in finance has been devoted to understanding the relationship between asset price volatility and trading volume; see, for example, Clark (1973), Tauchen and Pitts (1983), Andersen (1996), Bollerslev et al. (2018), and the many references therein. A common theme in this literature is the Mixture of Distribution Hypothesis (MDH), which postulates that the trading volume and price volatility are both driven by news arrival, and so, implies a positive relationship between volume and volatility. While the MDH is the leading theory for understanding the volume-volatility relationship, it does not force any specific functional form between these variables, which makes nonparametric procedures a natural choice for empirical study. In the first application, we apply the proposed nonparametric inferential tools to examine the volume-volatility relationship for the U.S. equity universe, with a particular focus on how this relationship may have changed after the breakout of the ongoing COVID-19 pandemic.

The dataset used in this empirical study consists of daily time series of Parkinson's volatility measure (Parkinson (1980)) and turnover for all stocks listed on NYSE, AMEX, and NASDAQ. Recall that Parkinson's volatility is defined as $Y_{it} = \log(h_{it}/l_{it})/\sqrt{4\log 2}$, where h_{it} and l_{it} are the intraday high and low prices of stock i on day t, and the stock's turnover X_{it} is defined as the number of traded shares divided by the total number of shares outstanding. Our post-COVID subsample ranges from February 24, 2020 to June 30, 2021, consisting of 342 trading days. The cutoff date, February 24, 2020, is when the U.S. government issued an official response to fight the pandemic. Meanwhile, we form the pre-COVID subsample using data from October 11, 2018 to February 21, 2020, so that it contains exactly the same number of trading days as the post-COVID sample. Our full sample thus contains 684 trading days in total and there are 7452 firms in the cross-section on an average day. To make the volatility and volume observations more comparable across firms, we consider the logarithm of each series and normalize it via its full-sample mean and standard deviation. As in the simulation study (see the online supplemental appendix), we further transform the (log-normalized) turnover onto the $[-1, 1]$ interval via the $x \mapsto 2\Phi(x) - 1$ transformation. The nonparametric procedure is then carried out using Legendre polynomials up to the sixth order.

Figure 1 plots the estimated conditional mean functions of volatility given turnover for the pre- and post-COVID subsamples. To gauge the sampling variability of these functional estimates, we also plot the associated 95% uniform confidence bands computed according to Algorithm 1. It is worth re-emphasizing that the uniform confidence band is designed to cover the entire conditional mean function simultaneously, and hence, allows us to make inferential statements regarding the whole function, rather than its value at a specific point. From the figure, we see that the confidence bands are reasonably tight, which highlights the advantage of harnessing the rich information from the panel dataset. The functional estimates reveal a clear positive relationship between volatility and turnover. Interestingly, the estimated curves appear to be steeper when the turnover is exceptionally high (roughly corresponding to the highest quintile). This finding suggests that, during extreme market conditions with "trading frenzy," the price impact of trades tends to be larger, which is likely due to an elevated level of asymmetric information among market participants during such circumstances (Kyle (1985)). In addition, we also note that the post-COVID estimate is higher than the pre-COVID estimate uniformly across all levels of trading activity. The non-overlapping confidence bands for the two estimated functions suggest that the observed difference between them is statistically highly significant (indeed, the formal test for their difference yields a p-value that is virtually zero). In view of the economic logic underlying

Figure 1: Volume-Volatility Relationship Before and After COVID-19

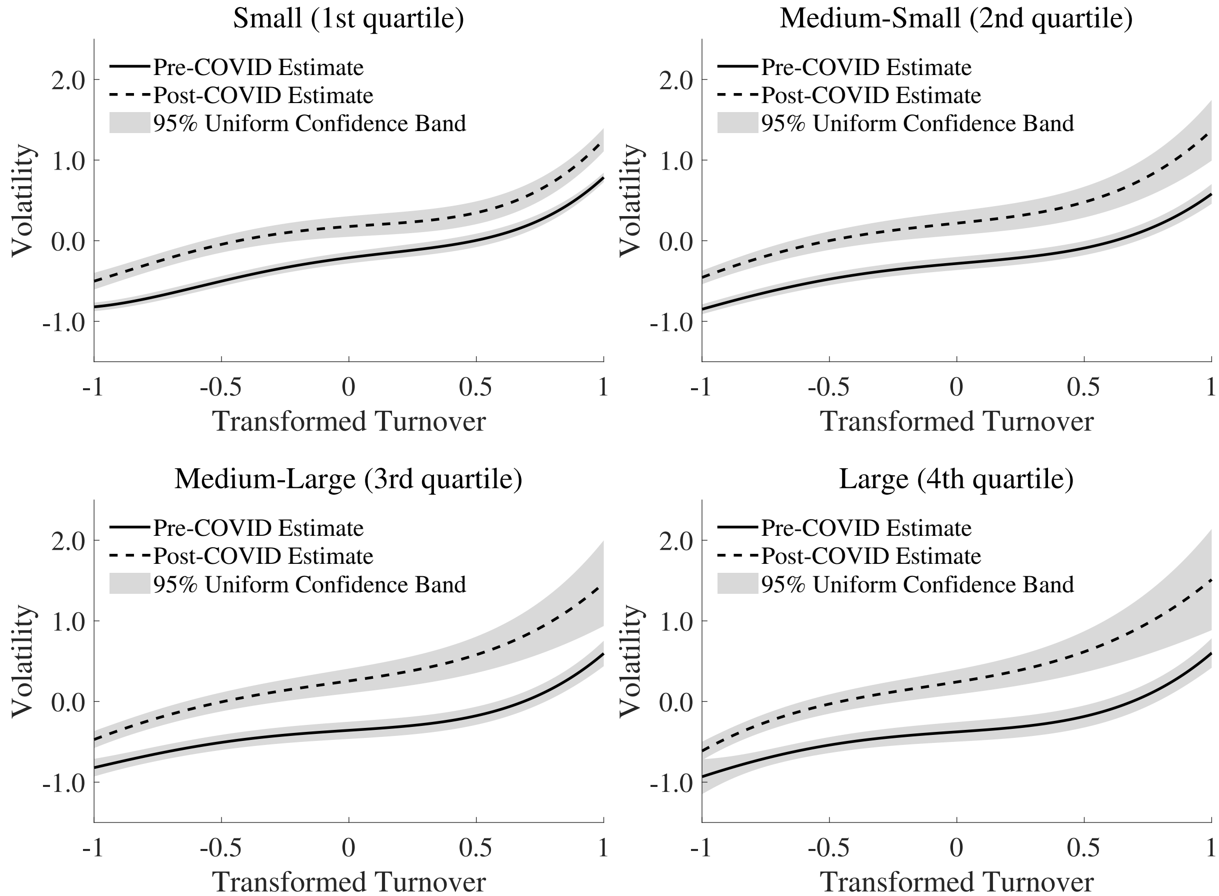


Note: The figure plots the nonparametric estimates of the conditional mean functions of the volatility given the (transformed) turnover for the pre- and post-COVID subsamples using all stocks. Each time series of volatility and turnover is log-normalized to have zero sample mean and unit standard deviation over the full sample period from October 11, 2018 to June 30, 2021. The transformed turnover is defined by transforming the normalized data via the $x \mapsto 2\Phi(x) - 1$ transformation so that it takes values on the $[-1, 1]$ interval. The nonparametric series estimation is implemented using Legendre polynomials up to the sixth order. The 95% uniform confidence bands are computed using Algorithm 1, where the Newey–West bandwidth is set to $\lfloor 0.75T^{1/3} \rfloor$.

the MDH, this result suggests that each "unit" of news during the pandemic period exerts more impact on stock price than it would do during normal times, even after conditioning on the same level of trading activity in a completely nonparametric fashion.

It is interesting to further investigate whether these findings may differ across small and large firms. To do so, we divide the stocks evenly into four quartile groups according to their market capitalizations. Figure 2 plots the nonparametric estimates and uniform confidence bands for these size-based groups. Remarkably, the results appear quite similar across all four groups. Hence, the findings seen from the full-sample analysis in Figure 1

Figure 2: Volume-Volatility Relationship for Small and Large Firms



Note: The figure plots the nonparametric estimate of the conditional mean function of volatility given the transformed turnover for each quartile of stocks sorted by their market capitalizations. The estimation is done separately for each size group following the same procedure as described in Figure 1.

are not driven by any particular size group. This observed homogeneity also confirms that pooling information among them, as we have done in Figure 1, is empirically sound.

3.2 Forecast rationality and information rigidity

In our second empirical application, we apply the proposed method to study the rationality of survey-based forecasts. Expectation surveys are routinely conducted among individuals, corporate officials, and professional researchers, and are widely used to gauge their forward-looking belief on the various aspects of socioeconomic life, business operation, and government policy. Arguably one of the most important surveys of this type is the Survey of Professional Forecasters (SPF), which collects institutional researchers' forecasts on leading macroeconomic indicators such as GDP, inflation, and unemployment. A large literature in macroeconomics has argued that these forecasts are not rational, in the sense that the ex post forecasting error can be partly predicted using a priori available information. An important mechanism for explaining this phenomenon is information friction. 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 forecasters and their average ex ante forecast revisions, and find empirical support for this hypothesis. 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.

It is instructive to briefly recall the empirical framework developed by Coibion and Gorodnichenko (2015). Let $Z_{k,t+h}$ denote the kth 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 $F_t Z_{k,t+h}$. 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:

$$Z_{k,t+h} - F_t Z_{k,t+h} = c + \beta \left(F_t Z_{k,t+h} - F_{t-1} Z_{k,t+h} \right) + e_{k,h,t}, \tag{3.1}$$

where the error term $e_{k,h,t}$ is mean-independent of the time-t information, and β 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 $\epsilon_{it} = e_{k,h,t}$. The specification in (3.1) may be rewritten more concisely as a linear panel regression

$$Y_{it} = c + \beta X_{it} + \epsilon_{it}. \tag{3.2}$$

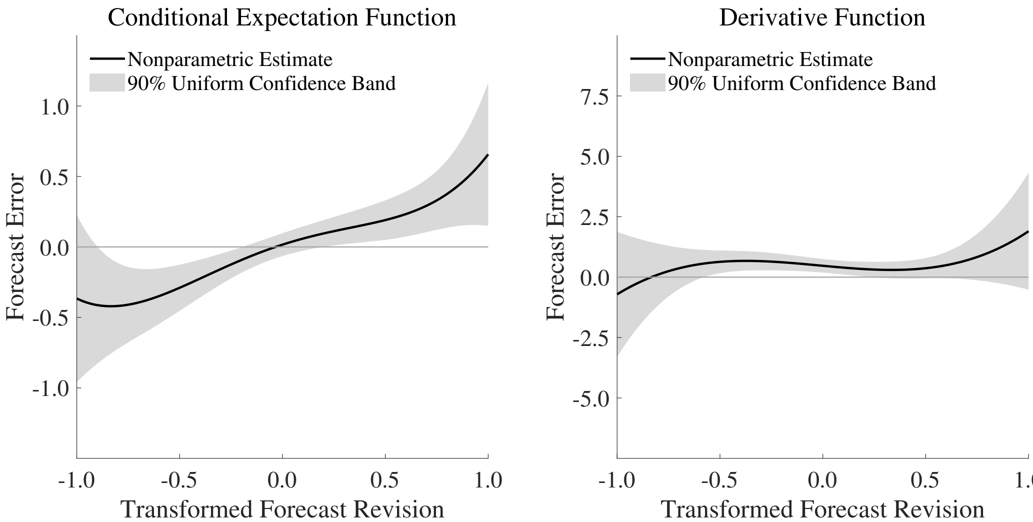
A natural generalization of (3.2) is the nonparametric regression in (2.1), namely, $Y_{it} = g(X_{it}) + \epsilon_{it}$. In the nonparametric context, information rigidity implies that $g(\cdot)$ is an increasing function. Compared to the baseline linear specification, the nonparametric model allows the marginal response (i.e., the derivative of $g(\cdot)$) 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.

We carry out the empirical analysis using the data from the SPF. For ease of comparison, we employ exactly the same data as in Coibion and Gorodnichenko (2015), 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 $T = 173$. 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. Like in Section 3.1, we normalize each time series with its sample mean and standard deviation, and transform the conditioning variable (i.e., forecast revision) onto the $[-1, 1]$ interval via $x \mapsto 2\Phi(x) - 1$. The series basis consists of Legendre polynomials up to the fifth order.

We now turn to the empirical results. On the left panel of Figure 3, we plot the nonparametric estimate of the conditional mean function of the forecast error given the transformed forecast revision, along with its 90% uniform confidence band computed according to Algorithm 1. 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, is consistent with the presence of information rigidity. To see this more clearly, on the right panel of Figure 3, 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 functional 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 β), 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 (3.2) 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;

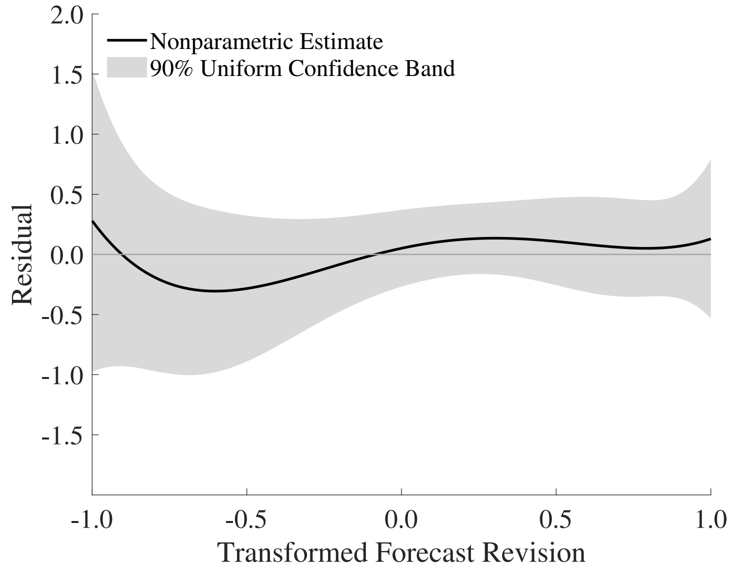
Figure 3: Nonparametric Estimation of Information Rigidity



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 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 4 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 (3.2). This suggests that the rigidity parameter β is likely constant across different levels of forecast revision and so Coibion and Gorodnichenko's (2015) linear specification (with constant β) is indeed adequate from this perspective.

Figure 4: 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 (3.2). 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 $\lfloor 0.75T^{1/3} \rfloor$.

4 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), $(T_n^{-1/2}\boldsymbol{H}_{n,t})_{t\geq 1}$ 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 $T_n^{-1/2}\boldsymbol{H}_{n,t}$. Therefore, applying Theorem 1 of Li and Liao (2020), we have

$$\|\boldsymbol{S}_n - \boldsymbol{\xi}_n\| = O_p(m_n^{1/2}r_n^{1/2} + (B_n m_n)^{1/3}) \tag{4.3}$$

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

$$\begin{aligned}
\sum_{t=1}^{T_n}\mathbb{E}[\|T_n^{-1/2}\boldsymbol{H}_{n,t}\|^3] &= \sum_{t=1}^{T_n}\mathbb{E}\left[\left(\sum_{j=1}^{m_n}(T_n^{-1/2}\boldsymbol{H}_{n,t}^{(j)})^2\right)^{3/2}\right] \\
&\leq T_n^{-3/2}m_n^{1/2}\sum_{t=1}^{T_n}\sum_{j=1}^{m_n}\mathbb{E}\left[|\boldsymbol{H}_{n,t}^{(j)}|^3\right] \\
&= O\left(T_n^{-1/2}m_n^{3/2}\right),
\end{aligned}$$

where the inequality is by Hölder's inequality and the last equality is by Assumption 1(iii). Plugging this estimate into (4.3), we readily deduce the assertion of Theorem 1. *Q.E.D.*

Proof of Theorem 2. Under condition (i), $(T_n^{-1/2}\boldsymbol{H}_{n,t})_{t\geq 1}$ is a mixingale satisfying Assumption 5 of Li and Liao (2020). Since $\bar{c}_n = O(1)$ 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

$$\|\boldsymbol{S}_n - \boldsymbol{\xi}_n\| = O_p(m_n^{1/2}T_n^{-1/2}) + O_p(m_n^{1/2}r_n^{1/2} + m_n^{5/6}T_n^{-1/6}) + O_p(m_nT_n^{-1/2} + m_n^{3/2}T_n^{-1}) \tag{4.4}$$

where $\boldsymbol{\xi}_n$ is an m_n-dimensional random vector with distribution $\mathcal{N}(0, \mathrm{Var}[\boldsymbol{S}_n])$. Since $m_n = o(T_n)$ by condition (iv), the assertion of Theorem 2 follows from (4.4). *Q.E.D.*

References

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance 51*(1), 169–204.

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.

Arellano, M. (1987). Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics 49*(4), 431–434.

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

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics 119*(1), 249–275.

Bollerslev, T., J. Li, and Y. Xue (2018). Volume, volatility, and public news announcements. *The Review of Economic Studies 85*(4), 2005–2041.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics* (1 ed.), Volume 6B, Chapter 76, pp. 5549–5632. Elsevier.

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

Chen, X. and X. Shen (1998). Sieve extremum estimates for weakly dependent data. *Econometrica 66*(2), 289–314.

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

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica 41*(1), 135–155.

Coibion, O. and Y. Gorodnichenko (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review 105*(8), 2644–2678.

Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

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

Driscoll, J. C. and A. C. Kraay (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics 80*(4), 549–560.

Eubank, R. L. and C. H. Spiegelman (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association 85*(410), 387–392.

Hansen, L. P. and R. J. Hodrick (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy 88*(5), 829–853.

Huang, J. Z. (1998). Projection estimation in multiple regression with application to functional anova models. *The Annals of Statistics 26*(1), 242 – 272.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics 31*(5), 1600 – 1635.

Jenish, N. (2012). Nonparametric spatial regression under near-epoch dependence. *Journal of Econometrics 167*(1), 224–239.

Kelejian, H. H. and I. R. Prucha (2007). HAC estimation in a spatial framework. *Journal of Econometrics 140*(1), 131–154.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1335.

Lee, J. and P. M. Robinson (2016). Series estimation under cross-sectional dependence. *Journal of Econometrics 190*(1), 1–17.

Li, J. and Z. Liao (2020). Uniform nonparametric inference for time series. *Journal of Econometrics 219*(1), 28–51.

Li, J., Z. Liao, and R. Quaedvlieg (2020). Conditional superior predictive ability. *Review of Economic Studies, forthcoming*.

Mankiw, N. G. and R. Reis (2002). Sticky information versus sticky prices: a proposal to replace the new keynesian phillips curve. *The Quarterly Journal of Economics 117*(4), 1295–1328.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications 9*(1), 141–2.

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

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business 53*(1), 61–65.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies 22*(1), 435–480.

Reis, R. (2006). Inattentive producers. *The Review of Economic Studies 73*(3), 793–821.

Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis 4*(3), 185–207.

Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics 165*(1), 5–19.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Tauchen, G. E. and M. Pitts (1983). The price variability-volume relationship on speculative markets. *Econometrica 51*(2), 485–505.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A (1961-2002) 26*(4), 359–372.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. In P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford (Eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, pp. 25–58. Princeton, United States: Princeton University Press.