The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–20

Uniform Nonparametric Inference for Spatially Dependent Panel Data: The xtnpsreg Command

Jia Li
Singapore Management University
jiali@smu.edu.sg

Zhipeng Liao
UCLA
zhipeng.liao@econ.ucla.edu

Wenyu Zhou
Zhejiang University
wenyuzhou@intl.zju.edu.cn

Abstract. In this article, we introduce a command, `xtnpsreg`, that implements a uniform nonparametric inference procedure for possibly unbalanced panel datasets with general forms of spatio-temporal dependence. We demonstrate how to apply this command in several use cases, including (i) the nonparametric estimation of conditional mean function and its marginal response; (ii) the construction of uniform confidence bands for these nonparametric functional parameters; (iii) specification tests for parametric model restrictions; and (iv) the estimation and uniform inference for functional coefficients in semi-nonparametric models.

Keywords: st????, `xtnpsreg`, series estimation, spatio-temporal dependence, uniform confidence band.

1 Introduction

Nonparametric regression methods provide a flexible way to study the relationship between variables. A popular approach is the series regression, which allows the user to approximate the unknown function with a "large" set of basis functions such as polynomials, splines, wavelet, etc. Conventional econometric theory (see, e.g., Andrews (1991) and Newey (1997)) allows one to conduct pointwise inference that is specific to the function's value at a given point. This may be unsatisfactory in practice because applied researchers are often interested in making inferential statements on the conditional mean function as a whole. The latter more demanding task requires uniform inference methods such as those developed by Belloni et al. (2015) and Li and Liao (2020) respectively for i.i.d. and serially dependent time-series data.

Meanwhile, panel datasets are widely used in various areas of empirical research. It is clearly of applied interest to conduct the aforementioned functional inference in the panel-data setting. An immediate benefit is that, by harnessing the richer information from both cross-sectional and time-series dimensions, one may obtain more accurate nonparametric estimates and draw sharper inference. This is a relevant consideration as the practical application of nonparametric methods is often hindered by a small sample size.

Much care is needed for performing reliable inference for panels, because this type of data often exhibits spatio-temporal dependence, namely, the observations may be mutually dependent on both cross-sectional and time-series dimensions, which has been emphasized by Bertrand et al. (2004) and Petersen (2009), among others. Not accounting for such dependence tends to result in an understatement of the sampling variability, leading the empiricist to mistakenly interpret "noise" as "signal."

A popular approach for dealing with spatio-temporal dependence is proposed by Driscoll and Kraay (1998) in the context of generalized method of moments. Driscoll–Kraay standard errors are robust to general forms of weak dependence in the time-series dimension and arbitrarily strong spatial dependence in the cross-sectional dimension. The underlying econometric theory requires "large T" asymptotics but does not restrict the dimensionality of the cross section. In Stata, `xtscc` implements Driscoll–Kraay standard errors for linear panel regressions. It is worth noting that in the degenerate case where the "panel" only contains a single time series, the Driscoll–Kraay standard error coincides with the classical Newey–West standard error (Newey and West (1987)); see [TS] **newey**.

In this paper, we propose a new command, `xtnpsreg`, which implements a panel (`xt`) nonparametric (`np`) series regression (`sreg`) and provides valid uniform functional inference that is robust to general forms of spatio-temporal dependence as considered in Driscoll and Kraay (1998). The underlying technical justification is detailed in a companion paper Li et al. (2021). The `xtnpsreg` command may be regarded as the nonparametric/functional version of `xtscc`. It is also related to the `tssreg` command developed by Li et al. (2020), which performs a similar task under the time-series setting. Roughly speaking, `xtnpsreg` extends `tssreg` in the same way as `xtscc` extends `newey`.

As we shall demonstrate in detail below, `xtnpsreg` may be conveniently used to carry out several types of nonparametric inferential tasks, including (i) the nonparametric estimation of a conditional mean function and its marginal response (i.e., the derivative function); (ii) the construction of uniform confidence bands for these functional parameters; (iii) nonparametric specification tests for parametric model restrictions; and (iv) the estimation and uniform inference for functional coefficients in semi-nonparametric models.

The remainder of this article is organized as follows. Section 2 provides some heuristics on the econometric/statistical theory underlying the proposed procedure. Section 3 documents the functionalities of the `xtnpsreg` command. Section 4 demonstrates how to use the new command to accomplish various nonparametric inferential tasks in an empirical example using data from the Federal Reserve Bank of Philadelphia Survey of Professional Forecasters.

2 Heuristics for the econometric procedure

In this section, we describe the econometric setting for the nonparametric regression problem, and provide some heuristics for the uniform functional inference procedure. For simplicity, we focus on the case with balanced panel in this discussion, while noting that unbalanced panels are accommodated by `xtnpsreg` as well.

2.1 Uniform functional inference for the conditional mean function

The baseline setting for the `xtnpsreg` command is the following nonparametric panel regression model:

$$Y_{it} = g(X_{it}) + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it}|X_{it}] = 0, \tag{1}$$

for $1 \leq i \leq N$ and $1 \leq t \leq T$. We assume that $T \to \infty$ but do not impose any restriction on N, that is, N may be fixed or grow to infinity. The object of interest is the conditional mean function $g(\cdot)$ that is implicitly defined as $g(x) = \mathbb{E}[Y_{it}|X_{it} = x]$. We focus on a single-equation setting with the dependent variable (*depvar*) Y_{it} being scalar-valued. In the current version of `xtnpsreg`, we also require the conditioning variable (*condvar*) X_{it} to be univariate. There are two reasons for this design choice. First, although multivariate conditioning is permitted in theory, the resulting nonparametric estimate tends to be imprecise because of the well-known "curse of dimensionality" in nonparametric analysis. For this reason, it advisable to single out a key explanatory variable X_{it} and allow it to enter the model nonparametrically. Second, the univariate conditioning also greatly simplifies the graphical presentation of the functional estimate and its confidence band, which is desirable for empirical discussions.

The main output of `xtnpsreg` consists of a nonparametric estimate for the conditional mean function $g(\cdot)$ and an associated uniform confidence band at a user-specified confidence level $1 - \alpha$. To be precise, the uniform confidence band is given by a pair of functional estimates $[L(\cdot), U(\cdot)]$ such that

$$\mathbb{P}(L(x) \leq g(x) \leq U(x) \text{ for all } x \in \mathcal{X}) \to 1 - \alpha, \quad \text{as} \quad T \to \infty. \tag{2}$$

In other words, the uniform confidence band covers the true conditional mean function simultaneously over the entire region $\mathcal{X}$ with approximately $1 - \alpha$ probability in large samples. By default, $\mathcal{X}$ is set to be the observed support of the conditioning variable. In certain applications, the user may want to take $\mathcal{X}$ as a subset of the observed support (so that the uniform nonparametric inference concentrates on a particular subregion of the conditioning space), which is allowed as an option.

The implementation of the statistical procedure proceeds as follows:

Step 1 (Nonparametric Estimation). The nonparametric estimator for $g(\cdot)$ is constructed by running a series regression. Specifically, let $P(X_{it}) = (p_1(X_{it}), \ldots, p_m(X_{it}))^\top$ denote a m-dimensional vector of approximating functions of X_{it}. Regressing Y_{it} on

$P(X_{it})$ yields the regression coefficient

$$\widehat{b} = \Big(\sum_{t=1}^{T}\sum_{i=1}^{N} P(X_{it})P(X_{it})^\top\Big)^{-1}\Big(\sum_{t=1}^{T}\sum_{i=1}^{N} P(X_{it})Y_{it}\Big), \tag{3}$$

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

$$\widehat{g}(\cdot) = P(\cdot)^\top \widehat{b}. \tag{4}$$

The current version of `xtnpsreg` employs Legendre polynomials to form the approximating functions $P(X_{it})$. Recall that the kth-order Legendre polynomial is given by $\frac{1}{2^k k!}\frac{d^k}{dx^k}(x^2-1)^k$. An important property of the Legendre polynomials is that they are orthogonal on the $[-1,1]$ interval with respect to the uniform distribution. This orthogonality property helps mitigate the multicollinearity among series terms, and hence, improves the numerical stability of the estimation procedure. Other types of orthogonal series basis may be adopted to serve the same purpose as well, and it might be interesting to incorporate them in a future version of `xtnpsreg`.

To better exploit the orthogonality property of Legendre polynomials, it is advisable to perform a preliminary transformation on the conditioning variable X_{it} so as to make it approximately uniformly distributed on the $[-1,1]$ interval. One way to achieve this is to consider some cumulative distribution function (CDF), say $F(\cdot)$, and transform X_{it} via $x \mapsto 2F(x)-1$. If $F(\cdot)$ is the CDF of X_{it}, the transformed variable will be exactly uniformly distributed on $[-1,1]$. In practice, setting $F(\cdot)$ as any reasonable approximation for the CDF of X_{it} can still achieve this goal to some extent, which will generally improve the numerical stability. By default, `xtnpsreg` employs the CDF of a normal distribution (calibrated to data) to carry out the transformation, which is an adequate choice provided that the distribution of the conditioning variable X_{it} roughly mimics a normal distribution. This default transformation may be disabled via the `method(none)` option, which allows the user to customize the transformation of the conditioning variable onto the $[-1,1]$ interval on their own.

Step 2 (Critical Value). The second step is to compute a critical value for a "functional t-statistic" that is defined as

$$\widehat{\tau} = \sup_{x\in\mathcal{X}} \frac{T^{1/2}|\widehat{g}(x)-g(x)|}{\widehat{\sigma}(x)}, \tag{5}$$

where $\widehat{\sigma}(x)$ is the estimated standard error for $\widehat{g}(x)$. Note that $\widehat{\tau}$ is simply the supremum of the pointwise t-statistics evaluated at different points over the conditioning space $\mathcal{X}$. The $\widehat{\sigma}(x)$ estimate is computed as

$$\widehat{\sigma}(x) = \sqrt{P(x)^\top \widehat{Q}^{-1}\widehat{A}\widehat{Q}^{-1}P(x)}, \tag{6}$$

where

$$\widehat{Q} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} P(X_{it})P(X_{it})^\top, \tag{7}$$

and $\widehat{A}$ is a "clustered" Newey–West estimator for the long-run variance-covariance matrix for the score vector $P(X_{it})\epsilon_{it}$ with the form (denoting $\widehat{\epsilon}_{it} = Y_{it} - \widehat{g}(X_{it})$)

$$\begin{cases} \widehat{A} = \displaystyle\sum_{s=-M_n}^{M_n} \frac{|M_n + 1 - s|}{M_n + 1} \widehat{\Gamma}_s, \text{ where} \\ \widehat{\Gamma}_s = \displaystyle\frac{1}{T} \sum_{t=\max\{1,1-s\}}^{\min\{T-s,T\}} \left(\frac{1}{N} \sum_{i=1}^{N} P(X_{it}) \widehat{\epsilon}_{it} \right) \left(\frac{1}{N} \sum_{i=1}^{N} P(X_{it+s}) \widehat{\epsilon}_{it+s} \right)^{\top}. \end{cases} \tag{8}$$

The user may specify the bandwidth parameter M_n via the `lag` option in `xtnpsreg` as in [TS] **newey**. We also note that the $\widehat{A}$ estimator is constructed in the same spirit as Driscoll and Kraay (1998) and it is robust with respect to general forms of spatio-temporal dependence.

The critical value of interest is an estimate for the $1-\alpha$ quantile of the sup-t statistic $\widehat{\tau}$. Li et al. (2021) show theoretically that the distribution of $\widehat{\tau}$ can be approximated in large sample by the conditional (given data) distribution of

$$\widehat{\tau}^* = \sup_{x \in \mathcal{X}} \frac{|P(x)^{\top} (\widehat{Q}^{-1} \widehat{A} \widehat{Q}^{-1})^{1/2} \mathcal{N}_m^*|}{\widehat{\sigma}(x)}, \tag{9}$$

where $\mathcal{N}_m^*$ is a generic m-dimensional standard normal random vector (which has the same dimensionality as $P(X_{it})$). To compute the critical value, we thus draw $\mathcal{N}_m^*$ from the standard normal distribution many times, and for each draw, compute $\widehat{\tau}^*$ over a discretized mesh of $\mathcal{X}$; we then set the critical value $cv_{1-\alpha}$ as the $1-\alpha$ empirical quantile of the simulated $\widehat{\tau}^*$.

In empirical applications, applied researchers are often interested in testing whether the conditioning variable may have any effect on the dependent variable, which in the present nonparametric setting amounts to testing the null hypothesis

$$H_0: \qquad g(x) \equiv E[Y_{it}|X_{it} = x] = 0, \qquad \text{for all } x \in \mathcal{X}. \tag{10}$$

We reject the null hypothesis at significance level α if the sup-t statistic $\widehat{\tau}$ (evaluated at $g(\cdot) = 0$) exceeds the critical value $cv_{1-\alpha}$. The test statistic, critical value, and the corresponding p-value are the default output of `xtnpsreg`.

Step 3 (Uniform Confidence Band). Finally, the $1-\alpha$ level two-sided uniform confidence band for $g(\cdot)$ is then given by

$$\mathrm{CB}_{1-\alpha}(\cdot) = [\widehat{g}(\cdot) - cv_{1-\alpha} T^{-1/2} \widehat{\sigma}(\cdot),\ \widehat{g}(\cdot) + cv_{1-\alpha} T^{-1/2} \widehat{\sigma}(\cdot)]. \tag{11}$$

The nonparametric functional estimate together with this confidence band can be displayed by activating the `plot` option; also see the `plotu` option for an alternative.

2.2 Marginal response

In linear regression models, the marginal effect of an explanatory variable on the dependent variable is completely summarized by its regression coefficient. For nonparametric regressions, the marginal response is captured by the derivative of the conditional mean function, denoted $\partial g(\cdot)$. By calling the `marginal` option, `xtnpsreg` computes a nonparametric functional estimate for $\partial g(\cdot)$ along with its $1-\alpha$ uniform confidence band. The implementation is carried out in the background as follows:

Step 1. Compute $\widehat{b}$, $\widehat{Q}$, and $\widehat{A}$ as in Section 2.1. Set

$$\partial \widehat{g}(\cdot) = \partial P(\cdot)^\top \widehat{b}, \qquad \tilde{\sigma}(x) = \sqrt{\partial P(x)^\top (\widehat{Q}^{-1}\widehat{A}\widehat{Q}^{-1})\partial P(x)}. \tag{12}$$

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

$$\widehat{\tau}'^* = \sup_{x\in\mathcal{X}} \frac{|\partial P(x)^\top (\widehat{Q}^{-1}\widehat{A}\widehat{Q}^{-1})^{1/2}\mathcal{N}_m^*|}{\tilde{\sigma}(x)}, \tag{13}$$

where the supremum can 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 $\widehat{\tau}'^*$.

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

2.3 Functional coefficient model

The aforementioned nonparametric uniform inference method can also be adapted to study linear regression models with functional coefficients. Specifically, consider the following specification

$$Y_{it} = c + \beta(X_{it})U_{it} + \epsilon_{it}, \quad \mathbb{E}[\epsilon_{it}|X_{it}, U_{it}] = 0, \tag{14}$$

where c is the intercept, U_{it} is a scalar-valued "base" explanatory variable (*basevar*), and $\beta(\cdot)$ is its functional coefficient modeled nonparametrically as a function of the conditioning variable X_{it}. The inferential target is the function $\beta(\cdot)$.

Note that the baseline setting described in Section 2.1 may be considered as a special case of (14) with $U_{it} = 1$ and $g(x) = c + \beta(x)$. The user may signify this more general functional coefficient setting by turning on the `funccoef` option in `xtnpsreg`. In this situation, the functional inference will concentrate on the functional coefficient $\beta(\cdot)$ without adding the intercept term c.

2.4 Semi-nonparametric model with linear control variables

For the settings discussed above, `xtnpsreg` also allows additional control variables to enter the model linearly. The generalized versions of the baseline nonparametric regression model (1) and the functional coefficient model (14) are given by, respectively,

$$Y_{it} = g(X_{it}) + Z_{it}^\top\gamma + \epsilon_{it}, \tag{15}$$

and

$$Y_{it} = c + \beta(X_{it})U_{it} + Z_{it}^{\top}\gamma + \epsilon_{it}, \tag{16}$$

where Z_{it} is a vector of control variables (*controlvar*). Under these settings, the focal point of the functional inference remains to be $g(\cdot)$ and $\beta(\cdot)$, respectively.

3 The `xtnpsreg` command

This section documents the syntax and functionalities of the `xtnpsreg` command. The command requires the `moremata` package, which may be installed in command line via `ssc install moremata`. The user also needs to declare the panel-data structure beforehand via `xts` *panelvar timevar*.

3.1 Syntax

The Stata syntax of the `xtnpsreg` command is as follows:

`xtnpsreg` *depvar condvar* [*basevar*] [*controlvar*] [*if*] [*in*] [`, lag(`*#*`) m(`*#*`)`
`method(`*transtype*`) confidencelevel(`*#*`) ngrid(`*#*`) mc(`*#*`) triml(`*#*`)`
`trimr(`*#*`) plot plotu scatter(`*#*`) funccoef marginal table excel`]

where *depvar* corresponds to the dependent variable Y_{it}, *condvar* denotes the nonparametric conditioning variable X_{it}, *basevar* is the univariate "base" explanatory variable U_{it} in the functional coefficient model described in Section 2.3, and *controlvar* contains a vector Z_{it} of linear control variables described in Section 2.4.

3.2 Options

`lag(`*#*`)` specifies the number of lags for computing the Newey–West estimator of the long-run variance-covariance matrix. The default is given by the integer part of $0.75T^{1/3}$, where T is the number of time periods.

`m(`*#*`)` specifies the number of Legendre polynomial terms used in the series estimation. The default is `m(6)`.

`method(`*transtype*`)` specifies the transformation implemented on the conditioning variable. 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. The approximating functions are Legendre polynomials of the transformed variable. The current version supports the following transformation methods, with `method(`*normal*`)` set to be the default.

- *none*: no transformation;

- *normal*: normal transformation $x \mapsto 2\Phi[(x - \overline{x})/\sigma] - 1$, where $\overline{x}$ and σ are the sample mean and standard deviation of x, and Φ is the cumulative distribution function of the standard normal distribution.

`confidencelevel(#)` specifies the confidence level (in percentage) of the uniform confidence band. The default is `confidencelevel(90)`.

`ngrid(#)` specifies the number of grid points used for discretizing the support of the transformed conditioning variable. The default is `ngrid(1000)`.

`mc(#)` specifies the number of Monte Carlo simulations used to compute the critical value. The default is `mc(5000)`.

`triml(#)` sets the left limit of the conditioning region $\mathcal{X}$ to be the # empirical quantile of *condvar*. The default is `triml(0)`.

`trimr(#)` sets the right limit of the conditioning region $\mathcal{X}$ to be the $1 - \#$ empirical quantile of *condvar*. The default is `trimr(0)`.

`plot` produces a plot of the nonparametric functional estimate and its uniform confidence band, in which the transformed conditioning variable is plotted on the horizontal axis.

`plotu` produces a plot of the nonparametric functional estimate and its uniform confidence band, in which the original conditioning variable is plotted on the horizontal axis.

`scatter(#)` adds a scatter plot of the data points, where # is a number between $[0, 100]$ that specifies the fraction of data points to be plotted.

`funccoef` signifies that the model of interest is a functional coefficient model with *basevar* as the base explanatory variable.

`marginal` implements the estimation of marginal response function.

`table` reports the estimated regression coefficients and standard errors in the series estimation.

`excel` generates an Excel file that contains the requisite information for plotting the functional estimate and the associated uniform confidence band.

3.3 Stored results

The `xtnpsreg` command stores the following results in `e()`:

Scalars

`e(N)`	number of cross-sectional units
`e(T)`	number of time periods
`e(supt)`	sup-t statistic
`e(cv)`	critical value for the sup-t test
`e(df_r)`	residual degrees of freedom

Macros

`e(depvar)`	name of the dependent variable
`e(condvar)`	name of the conditioning variable
`e(method)`	transformation method
`e(cmd)`	`xtnpsreg`

Matrices

`e(b)`	regression coefficients in series estimation
`e(se)`	standard errors of regression coefficients in series estimation
`e(V)`	variance-covariance matrix of the regression coefficients
`e(ygrid)`	functional estimate
`e(xgrid)`	grid points of the conditioning variable
`e(sigma)`	estimate of standard error function

4 An empirical illustration

In this section, we demonstrate `xtnpsreg`'s main usage in an example built on the empirical analysis of Coibion and Gorodnichenko (2015) and Li et al. (2021). The dataset and implementation code are provided in the online supplement accompanying this paper.

4.1 Data description and empirical motivation

The `spf.dta` dataset is constructed from the Survey of Professional Forecasters. It contains quarterly time series of ex post forecast errors (`fe`) and ex ante forecast revisions (`fr`) averaged among forecasters from 1969 to 2014 for five macroeconomic variables, including GDP price deflator, real GDP, industrial production, housing starts, and unemployment rate, over four forecast horizons. We treat the data for each variable-horizon pair as an individual time series. Merging them yields a panel dataset with $N = 20$ and $T = 173$.

Since the time series of forecast errors and forecast revisions have quite different scales across variables, we first normalize them separately so that each time series is averaged at zero with unit standard deviation. The resulting normalized forecast error and forecast revision are stored as `fe_norm` and `fr_norm`. The normalization is implemented as follows.

```
. *** Load Data ***
. use "spf.dta", clear

. xtset id_N id_date
       panel variable:  id_N (strongly balanced)
        time variable:  id_date, 1 to 173
                delta:  1 unit
```

```
.
. *** Data Normalization ***
. by id_N: egen fe_id_mean = mean(fe)
. by id_N: egen fe_id_sd   = sd(fe)
. by id_N: gen  fe_norm  = (fe-fe_id_mean)/fe_id_sd
.
. by id_N: egen fr_id_mean = mean(fr)
. by id_N: egen fr_id_sd   = sd(fr)
. by id_N: gen  fr_norm  = (fr-fr_id_mean)/fr_id_sd
```

The economic motivation for studying the relationship between ex post forecast error and ex ante forecast revision is to examine whether the professional forecasters are collectively rational. Under the rational-expectation hypothesis, the forecast errors are entirely unanticipated, and so, their conditional expectation given any a priori known information (including the forecast revision) should be zero. Meanwhile, a large literature in macroeconomics argues that the full rationality benchmark may break down due to information stickiness, which in turn implies a positive relationship between forecast error and forecast revision as shown in Coibion and Gorodnichenko (2015). We may assess the empirical plausibility of these alternative theoretical predictions by nonparametrically regressing the forecast error on the forecast revision. Below, we demonstrate how to use the `xtnpsreg` command to implement the nonparametric estimation and the related functional inference.

4.2 The basic use case of `xtnpsreg`

The main, and most basic, use of `xtnpsreg` is to nonparametrically estimate the conditional mean function and plot the functional estimate together with its uniform confidence band. As a first illustration, we nonparametrically regress the (normalized) forecast error `fe_norm` on the forecast revision `fr_norm` as follows.

```
. xtnpsreg fe_norm fr_norm, plot
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	4.2721	2.5412	0.000

The output table reports the sup-t statistic, critical value, and p-value for testing the null hypothesis that the conditional mean function is identically zero in a uniform sense. The default significance level is $\alpha = 10\%$, which may be changed via the `confidencelevel(#)` option (e.g., `confidencelevel(95)` corresponds to $\alpha = 5\%$). The table above shows that the sup-t statistic is notably greater than the critical value, indicating a strong rejection of the null hypothesis. Indeed, the virtually zero p-value suggests that the null hypothesis is also rejected at, say, the 1% significance level. This finding implies that the forecasts are not fully rational.

Figure 1 plots the estimated conditional mean function and the associated 90%

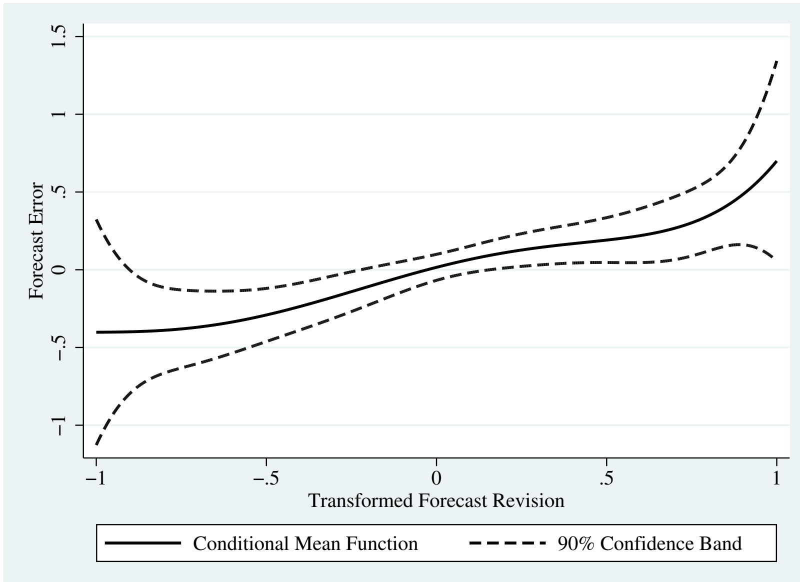

Figure 1: Default output of estimated conditional mean function and the 90% uniform confidence band.

uniform confidence band. By default, `xtnpsreg` transforms the conditioning variable `fr_norm` onto the $[-1, 1]$ via the $x \mapsto 2\Phi(x) - 1$ transformation, where Φ denotes the normal CDF calibrated using the sample mean and variance of the conditioning variable. The fact that the confidence band does not always cover the zero horizontal line means that the conditional mean function is statistically different from zero as a whole, which, needless to say, is consistent with the aforementioned testing result. The plot also reveals that the conditional mean of the forecast error is an increasing function in the forecast revision, and so, provides support for theoretical predictions from information-rigidity models. The requisite information for generating Figure 1 may be exported to a spreadsheet by calling the `excel` option.

4.3 Robustness checks with respect to tuning parameters

The proposed nonparametric econometric method mainly involves two tuning parameters. One is the number of series terms `m(#)`. The default specification is `m(6)`, which corresponds to a fifth-order Legendre polynomial. The other is the bandwidth parameter `lag(#)` stemming from the computation of the Newey–West type standard error, which is set to be the integer part of $0.75T^{1/3}$ by default. In theory, one should employ a larger number of series terms for larger samples, and use more lags if the data exhibits stronger serial dependence on the t dimension. But it is difficult in practice to pin down these choices "optimally." It is thus useful to check the robustness of empirical findings with respect to these choices.

As a concrete demonstration, we repeat the nonparametric estimation with different numbers of series terms (4, 6, 8, 10) by modifying the `m(#)` option as follows.

```
. xtnpsreg fe_norm fr_norm, m(4) plot
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	4.9230	2.3875	0.000

```
. xtnpsreg fe_norm fr_norm, m(6) plot
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	4.2721	2.5358	0.001

```
. xtnpsreg fe_norm fr_norm, m(8) plot
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	5.2039	2.5882	0.000

```
. xtnpsreg fe_norm fr_norm, m(10) plot
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	5.4878	2.6891	0.000

We may also check the effect of the Newey–West lag parameter by modifying the `lag(#)` option as follows.

```
. xtnpsreg fe_norm fr_norm, lag(2) plot
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	4.5055	2.5561	0.000

```
. xtnpsreg fe_norm fr_norm, lag(4) plot
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	4.2721	2.5218	0.001

```
. xtnpsreg fe_norm fr_norm, lag(6) plot
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	4.1318	2.5026	0.000

```
. xtnpsreg fe_norm fr_norm, lag(8) plot
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	4.0872	2.5225	0.001

From these tables and the plots in Figures 2 and 3, we see that the empirical results

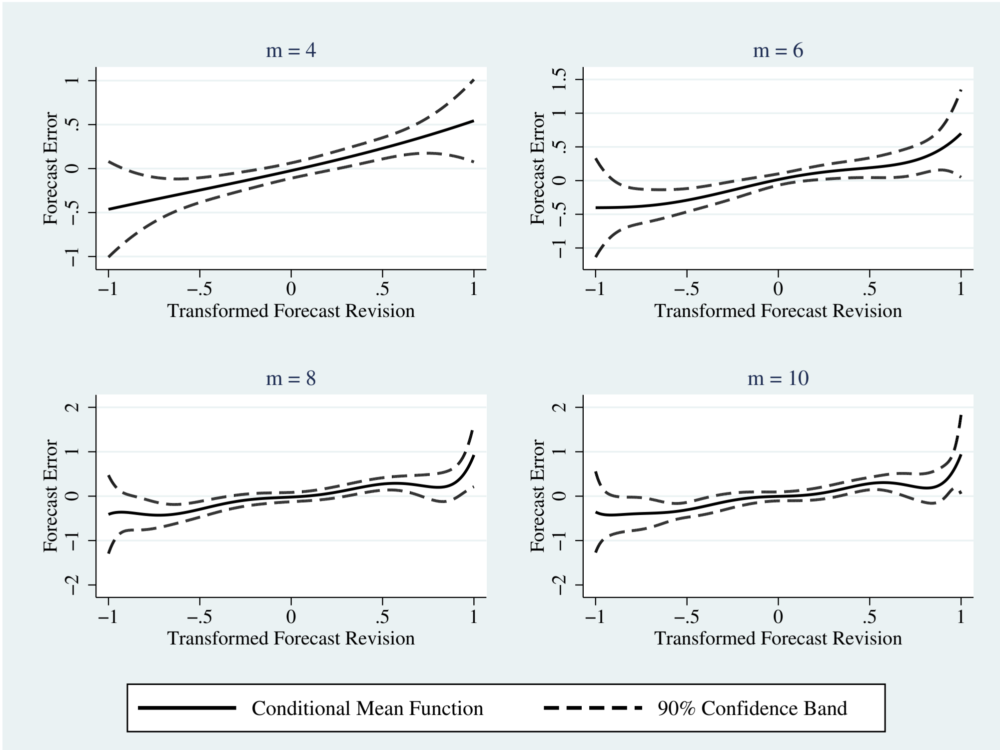

Figure 2: Nonparametric estimates with different numbers of series terms.

are fairly robust with respect to different choices of the number of series terms and/or the number of Newey–West lags.

4.4 Nonparametric inference for marginal response

Besides the conditional mean function $g(\cdot)$ itself, applied researchers may be interested in estimating the marginal response, defined as the derivative function $\partial g(\cdot)$. The nonparametric estimate and the associated uniform confidence band can be computed via `xtnpsreg` by calling the `marginal` option as follows.

```
. xtnpsreg fe_norm fr_norm, plot marginal
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	4.3819	2.5312	0.000

In this table, the sup-t statistic, critical value, and p-value pertain to testing the null hypothesis that the derivative function $\partial g(\cdot)$ is identically zero. The results sug-

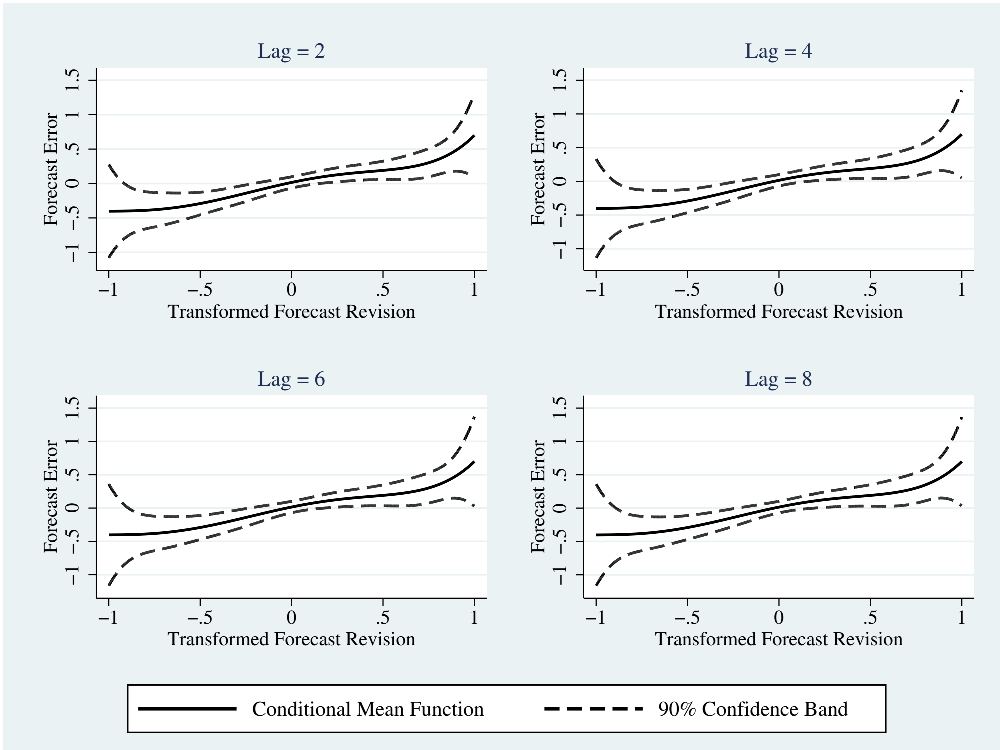

Figure 3: Nonparametric estimates with different numbers of Newey–West lags.

gest that the marginal response function as a whole is statistically different from zero. Figure 4 plots the nonparametric estimate of the marginal response and its 90% uniform confidence band. From the figure, we see that the estimated marginal response is nonnegative over the conditioning space, which is consistent with the previous observation that the conditional mean function of forecast error is increasing in the amount of forecast revision as predicted by information-rigidity models.

4.5 Specification tests

The uniform nonparametric inference method may also be used to conduct nonparametric specification tests against parametric model restrictions. For instance, we may formally test whether a linear specification is sufficient to describe the relationship between forecast error and forecast revision. To do so, we first run an ordinary least squares regression of forecast error on forecast revision and obtain the residual as follows.

```
. reg fe_norm fr_norm
      Source |       SS           df       MS      Number of obs   =     3,460
-------------+----------------------------------   F(1, 3458)      =    207.14
             |
```

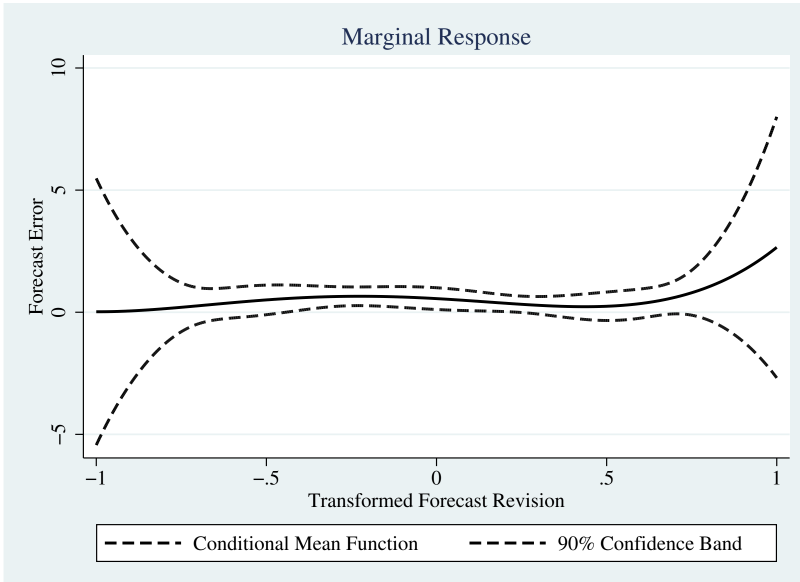

Figure 4: Nonparametric estimate and uniform confidence band of the derivative function.

Model	194.412734	1	194.412734	Prob > F =	0.0000
Residual	3245.58726	3,458	.938573529	R-squared =	0.0565
				Adj R-squared =	0.0562
Total	3440	3,459	.994507082	Root MSE =	.9688

fe_norm	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
fr_norm	.2377295	.0165179	14.39	0.000	.2053437	.2701154
_cons	4.32e-10	.0164701	0.00	1.000	-.0322921	.0322921

```
. predict fresidual, residuals
```

If the linear specification for the conditional mean function is correct, the conditional expectation of the residual given the conditioning variable should be zero. To test this formally, we use `xtnpsreg` to implement the nonparametric regression as follows.

```
. xtnpsreg fresidual fr_norm
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	2.1216	2.5186	0.239

From the table, we see that the null hypothesis of correct specification cannot be rejected at the 10% level, which suggests that the linear specification is in fact compatible with the observed data.

It is worth noting that the residual obtained from the linear regression are "generated variables" in that they are noisy approximations for the unobserved disturbance terms. That noted, it can be shown theoretically (see Li and Liao (2020)) that this approximation error is asymptotically negligible for the nonparametric specification test. The intuition is that, in large samples, estimation errors in the linear regression coefficients shrink to zero at a faster rate than the statistical error in the nonparametric test. This theoretical intuition works better when the number of series terms is relatively large. It is thus advisable to check the robustness of the empirical finding by increasing `m(#)` as shown in the following implementation.

```
. xtnpsreg fresidual fr_norm, m(10)
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	2.4977	2.7045	0.164

4.6 Semi-nonparametric setting with linear control variables

We next describe how to use `xtnpsreg` in the partial linear model (15) where a vector Z_{it} of control variables enters the specification linearly. For this illustration, we randomly generate two control variables `Z1` and `Z2`, and feed them to `xtnpsreg` as *controlvar*. We also turn on the `table` option to display the estimated coefficients for all series terms and control variables.

```
. gen Z1 = rnormal()
. gen Z2 = rnormal()
. xtnpsreg fe_norm fr_norm Z1 Z2, table
Number of obs         =      3460
Newey-West maximum lag =        4
```

	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
p_1(fr_norm)	-.0101484	.0331531	-0.31	0.760	-.0756074	.0553105
p_2(fr_norm)	.279338	.0640112	4.36	0.000	.1529513	.4057247
p_3(fr_norm)	.0172467	.027071	0.64	0.525	-.0362035	.070697
p_4(fr_norm)	.0059171	.0354302	0.17	0.868	-.0640379	.0758721
p_5(fr_norm)	.0398964	.0232038	1.72	0.087	-.0059183	.0857112
p_6(fr_norm)	.0157862	.0301797	0.52	0.602	-.0438019	.0753744
Z1	-.0061957	.0149366	-0.41	0.679	-.0356873	.0232958
Z1	.0168962	.0133845	1.26	0.209	-.0095308	.0433232

Transformation:	sup-t	10% critical value	P>|t|
Normal	4.2755	2.5429	0.000

As expected, the coefficients of `Z1` and `Z2` are both close to zero and statistically insignificant, as they are simply irrelevant for the data generating process. Meanwhile, the sup-t statistic and its critical value remain to be very similar to those seen in Section

4.2.

4.7 Uniform inference for functional coefficients

We now demonstrate how to use `xtnpsreg` to conduct inference in the functional coefficient model (14) and its generalized version (16). In this illustration, the dependent variable Y_{it} remains to be the forecast error. But we now set the forecast revision as the base explanatory variable U_{it}, and its marginal effect is given by the function $\beta(\cdot)$ of a new conditioning variable X_{it}. We take X_{it} as 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. In this example, X_{it} happens to be a univariate time series not depending on i; this is permitted, but not required, by the estimation procedure.

The time series of the log market volatility is stored as `mktvol` in the dataset `volatility.dta` provided in the online supplemental material. As a preliminary preparation, we need to convert the univariate volatility series into a panel by merging the original panel dataset `spf.dta` with the new one as follows.

```
. *** Merge data ***
. merge m:1 id_date using volatility
    Result                           # of obs.
    ─────────────────────────────────────────
    not matched                              0
    matched                              3,460  (_merge==3)
    ─────────────────────────────────────────
```

The functional coefficient model (14) is estimated by turning on the `funccoef` option in `xtnpsreg` as follows.

```
. xtnpsreg fe mktvol fr, plot funccoef
```

Transformation:	sup-t	10% critical value	P>|t|
Normal	3.6591	2.5624	0.004

It is instructive to clarify the syntax of this command. Here, `fe` and `mktvol` are parsed as *depvar* and *condvar*, respectively. With the `funccoef` option turned on, the variable that immediately follows *condvar* (i.e., `fr`) is parsed as *basevar*, and the remaining list of variables, if there is any, is parsed as *controlvar*. Without turning on `funccoef`, `xtnpsreg` would instead parse all variables following *condvar* as *controlvar* as described in Section 4.6 above. The command above also generates a plot for the nonparametric estimate of the functional coefficient $\beta(\cdot)$ and its uniform confidence band, as shown in Figure 5 below.

As in the baseline setting, `xtnpsreg` transforms the conditioning variable using the normal CDF to the $[-1,1]$ interval because the `method(`*normal*`)` option is active by default. The functional estimates in Figure 5 are plotted under the transformed scale, which explains the $[-1,1]$ domain on the horizontal axis. The user may also obtain

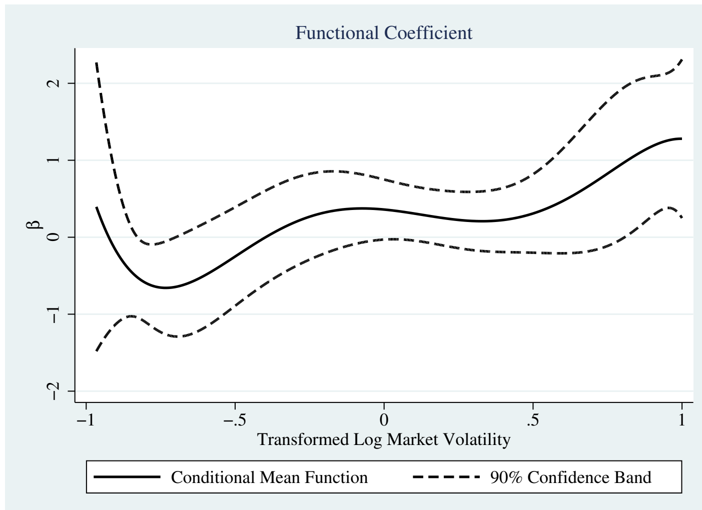

Figure 5: Nonparametric estimate and uniform confidence band of the functional coefficient $\beta(\cdot)$ plotted on the transformed scale using the `plot` option.

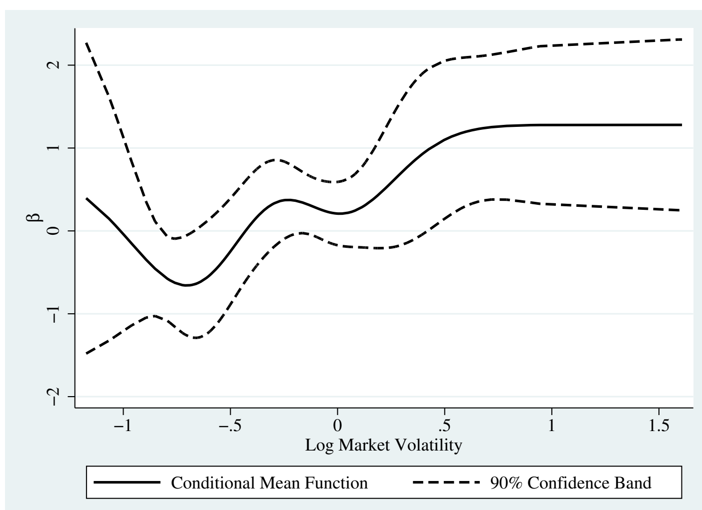

Figure 6: Nonparametric estimate and uniform confidence band of the functional coefficient $\beta(\cdot)$ plotted on the original scale using the `plotu` option.

plots on the original untransformed scale of *condvar* by replacing `plot` with `plotu`, as shown in Figure 6.

4.8 Implementation for "large N small T" panels via index swapping

As discussed in Section 2, the proposed method relies on a "large T" setting but does not restrict the cross-sectional dimension, which may be fixed or divergent. Correspondingly, the underlying theory also requires that the dependence along the time-series dimension is weak, whereas the cross-sectional dependence is allowed to be arbitrarily strong.

It is important to note that whether one labels i or t as "individual" or "time" is completely inconsequential. For all econometric purposes, what matters is that i indexes the dimension with possibly strong dependence and arbitrary sample size and t indexes the dimension with weak dependence (with independence being a special case) and large sample size.

Therefore, by swapping the role of i and t (so that i and t become the time and cross-sectional indexes, respectively), we may also apply `xtnpsreg` in short panels with a large number of independent cross-sectional units. Due to the lack of dependence in the new t dimension, the Newey–West lag should be set to `lag(0)`. Arbitrarily strong serial dependence is accommodated as time is now indexed by i.

5 References

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

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato. 2015. Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results. *Journal of Econometrics* 186(2): 345 – 366.

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.

Coibion, O., and Y. Gorodnichenko. 2015. Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review* 105(8): 2644–78.

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

Li, J., and Z. Liao. 2020. Uniform nonparametric inference for time series. *Journal of Econometrics* 219(1): 38–51.

Li, J., Z. Liao, and M. Gao. 2020. Uniform nonparametric inference for time series using Stata. *The Stata Journal* 20(3): 706–720.

Li, J., Z. Liao, and W. Zhou. 2021. Uniform Nonparametric Inference for Spatially Dependent Panel Data. *Working Paper* .

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.

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

About the authors

Jia Li is the Lee Kong Chian Professor of Economics at the School of Economics, Singapore Management University.

Zhipeng Liao is an associate professor of economics at the Department of Economics, UCLA.

Wenyu Zhou is an assistant professor at the International Business School, Zhejiang University.