Journal of Econometrics 188 (2015) 346–362

Contents lists available at ScienceDirect

Journal of Econometrics

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

# Panel nonparametric regression with fixed effects

Jungyoon Lee [a], Peter M. Robinson [b,*]

[a] New College of the Humanities, 19 Bedford Square, London WC1B 3HH, UK
[b] London School of Economics, Houghton Street, London WC2A 2AE, UK

ARTICLE INFO

*Article history:*
Available online 11 March 2015

*JEL classification:*
C13
C14
C23

*Keywords:*
Panel data
Nonparametric regression
Cross-sectional dependence
Generalized least squares
Optimal bandwidth

ABSTRACT

Nonparametric regression is developed for data with both a temporal and a cross-sectional dimension. The model includes additive, unknown, individual-specific components and allows also for cross-sectional and temporal dependence and conditional heteroscedasticity. A simple nonparametric estimate is shown to be dominated by a GLS-type one. Asymptotically optimal bandwidth choices are justified for both estimates. Feasible optimal bandwidths, and feasible optimal regression estimates, are also asymptotically justified. Finite sample performance is examined in a Monte Carlo study.



## 1. Introduction

The advantages of panel data have been exploited in many econometric model settings, following the early and influential contributions of Cheng Hsiao (see e.g. Hsiao, 1986). Much of the literature stresses parametric regression and/or time trending effects, alongside unknown individual effects. Nonparametric models lessen the risk of misspecification and can be useful in relatively large data sets, and have already featured in panel settings. Ruckstuhl et al. (2000) asymptotically justified nonparametric regression estimation when time series length $T$ increases and cross-sectional size $N$ is fixed, and there is no cross-sectional dependence. When allowing possible dependence in either time or cross-sectional dimension (or both), the question of efficiency improvement via utilizing correlation structure arises naturally. Carroll et al. (2003) and Wang (2003) explored the possibility of efficiency gain in nonparametric regression estimation by exploiting temporal correlation but under cross-sectional independence; Henderson et al. (2008) estimated nonparametric and partly linear regressions with additive individual effects; Evdokimov (2010) considered identification and estimation in nonparametric regression with nonadditive individual effects; Li et al. (2011) studied nonparametric time-varying coefficient panel data models; Hoderlein et al. (2011) considered nonparametric binary choice models with fixed effects; Koerber et al. (forthcoming) dealt with nonparametric regression model with individual and time fixed effects, where the regression function can vary across individuals; under temporal independence, Robinson (2012) efficiently estimated a nonparametric trend in the presence of possible cross-sectional dependence.

The present paper considers efficiency improvement where the nonparametric regression is a function of a possibly vector-valued observable stationary sequence that is common to all cross-sectional units, addressing similar issues as Robinson (2012). As in that paper, $T$ is assumed large relative to $N$, as can be relevant when the cross-sectional units are large entities such as countries/regions or firms. Disturbances may exhibit cross-sectional dependence due to spillovers, competition, or global shocks, and such dependence, of a general and essentially nonparametric nature, is allowed. We describe an observable array $Y_{it}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$, by

$$Y_{it} = \lambda_i + m(Z_t) + U_{it}, \quad i = 1, \ldots, N, \; t = 1, \ldots, T, \tag{1}$$

where the $\lambda_i$ are unknown nonstochastic individual fixed effects, $Z_t$ is a $q$-dimensional vector of time-varying stochastic regressors that are common to individuals, $m$ is a nonparametric function, and $U_{it}$ is an unobservable zero-mean array. The common trend model of Robinson (2012) replaced $Z_t$ by the deterministic argument $t/T$. He showed how to improve on simple estimates of $m$ by generalized least squares (GLS) ones using estimates of the cross-sectional variance matrix of $U_{it}$. Employing instead a stochastic $Z_t$ requires

* Corresponding author. Tel.: +44 20 7955 7516; fax: +44 20 7955 6592.
*E-mail addresses:* jungyoon.lee@nchum.org (J. Lee), p.m.robinson@lse.ac.uk (P.M. Robinson).

http://dx.doi.org/10.1016/j.jeconom.2015.03.004
0304-4076/

somewhat different methodology and substantially different asymptotic theory and admits the possibility of conditional heteroscedasticity of $U_{it}$. Furthermore, though he discussed implications of serial dependence in $U_{it}$, Robinson (2012) assumed temporal independence; we allow $U_{it}$ to be a weakly dependent stationary process with nonparametric autocorrelation. In addition, whereas Robinson (2012) focused on mean squared error (MSE) properties, we also establish asymptotic normality of estimates of $m$. Throughout, asymptotic theory is with respect to $T \to \infty$, with either $N \to \infty$ slowly relative to $T$, or $N$ fixed. When $N$ is fixed, the model can be considered a nonparametric seemingly unrelated regression. We discuss restrictions on the rate at which $N$ is allowed to grow with $T$ when relevant, see later in Theorems 6 and 7.

While (1) is of practical interest in itself, it can be more broadly motivated from a semiparametric model involving also time-varying, individual-specific regressors. For example, if $Y_{it}$ denotes a house price index of Eurozone countries, $Z_t$ the interest rate set by the European Central Bank, and $X_{it}$ country-specific covariates (such as GDP, inflation and stock market index), consider the partly linear specification

$$Y_{it} = \lambda_i + X_{it}'\gamma + m(Z_t) + U_{it}. \tag{2}$$

For a generic array $\{\varsigma_{it}\}$, $t = 1, \dots, T$, $i = 1, \dots, N$, denote for temporal, cross-sectional, and overall averages

$$\bar{\varsigma}_{At} = \frac{1}{N}\sum_{j=1}^{N} \varsigma_{jt}, \qquad \bar{\varsigma}_{iA} = \frac{1}{T}\sum_{s=1}^{T} \varsigma_{is}, \qquad \bar{\varsigma}_{AA} = \frac{1}{TN}\sum_{j=1}^{N}\sum_{s=1}^{T} \varsigma_{js}, \tag{3}$$

and put $\check{\varsigma}_{it} = \varsigma_{it} - \bar{\varsigma}_{At} - \bar{\varsigma}_{iA} + \bar{\varsigma}_{AA}$. We can thence transform (2) to

$$\check{Y}_{it} = \check{X}_{it}'\gamma + \check{U}_{it}. \tag{4}$$

Denoting by $\widehat{\gamma}$ an estimate of $\gamma$ obtained from (4) by, for example, least squares, at a rate that can be faster under suitable conditions than the nonparametric rate for estimation of $m$ (see e.g. Moon and Phillips, 1999) the methods developed in the paper should be justifiable with $Y_{it}$ in (1) replaced by $Y_{it} - X_{it}'\widehat{\gamma}$.

The plan of the paper is as follows. The following section introduces a simple kernel estimate of $m$ and presents its asymptotic MSE and the consequent optimal choice of bandwidth, and establishes its asymptotic normality. Section 3 presents infeasible generalized least squares (GLS) estimate of $m$ using the unknown cross-sectional covariance matrix of $U_{it}$, with asymptotic properties. In Section 4 feasible GLS estimate of $m$ is justified. Section 5 presents a small Monte Carlo study of finite sample performance. Proofs of theorems are provided in Appendix A, while Appendix B contains some useful lemmas, of which Lemma 6 constitutes an additional contribution in offering a decomposition of U-statistics of order up to 4, under serial dependence.

## 2. Simple non-parametric regression estimation

We can write (1) in $N$-dimensional vector form as

$$Y_{\cdot t} = \lambda + m(Z_t)1_N + U_{\cdot t}, \quad t = 1, \dots, T, \tag{5}$$

where $Y_{\cdot t} = (Y_{1t}, \dots, Y_{Nt})'$, $\lambda = (\lambda_1, \dots, \lambda_N)'$, $1_N = (1, \dots, 1)'$, $U_{\cdot t} = (U_{1t}, \dots, U_{Nt})'$, the prime denoting transposition. In (1), $\lambda_i$ and $m$ are identified only up to a location shift. As in Robinson (2012), the (arbitrary) restriction

$$\sum_{i=1}^{N} \lambda_i = 0 \tag{6}$$

identifies $m$ up to vertical shift and leads to

$$\bar{Y}_{At} = m(Z_t) + \bar{U}_{At}. \tag{7}$$

(See Lin et al., 2014 for an alternative identification condition when the $\lambda_i$ are stochastic.) From (7), we can nonparametrically estimate $m$ using the time series data $(\bar{Y}_{At}, Z_t')$. We employ the Nadaraya–Watson (NW) estimate

$$\tilde{m}(z) = \frac{\tilde{m}_n(z)}{\tilde{m}_d(z)},$$

where the numerator and denominator are given by

$$\tilde{m}_n(z) = \sum_{t=1}^{T} K\left(\frac{Z_t - z}{a}\right)\bar{Y}_{At}, \qquad \tilde{m}_d(z) = \sum_{t=1}^{T} K\left(\frac{Z_t - z}{a}\right),$$

$a$ is a positive bandwidth, and

$$K(u) = \prod_{j=1}^{q} k(u_j), \quad u = (u_1, u_2, \dots, u_q)', \tag{8}$$

where $k$ is a univariate kernel function. More general, non-product choices of $K$, and/or a more general diagonal or non-diagonal matrix-valued bandwidth, could be employed in practice but (8) with a single scalar bandwidth affords relatively simple conditions. Let $\mathcal{K}_\ell$, $\ell \geq 1$, denote the class of even $k$ satisfying

$$\int_{\mathbb{R}} k(u)du = 1, \qquad \int_{\mathbb{R}} u^i k(u)du = 0, \quad i = 1, \dots, \ell - 1,$$
$$0 < \int_{\mathbb{R}} u^\ell k(u)du < \infty, \qquad \sup_u (1 + |u|^{\ell+1})|k(u)| < \infty.$$

We introduce regularity conditions on $Z_t$, $U_{it}$ similar to those employed in the pure time series case by Robinson (1983) and in subsequent references.

**Assumption 1.** For all sufficiently large $i$, $(Z_t', U_{1t}, \dots, U_{it})'$ is jointly stationary and $\alpha$-mixing with mixing coefficient $\alpha_i(j)$, and for some $\mu > 2$, $\alpha(j) = \overline{\lim}_{i\to\infty}\alpha_i(j)$ satisfies

$$\sum_{j=n}^{\infty} \alpha^{1-2/\mu}(j) = o(n^{-1}), \quad \text{as } n \to \infty.$$

Assumption 1 is from Robinson (1983) and imposes a mild restriction on the rate of decay in the strong mixing coefficient.

**Assumption 2.** For all $i \geq 1$, $t \geq 1$, $E(U_{it}|Z_t) = 0$ almost surely (a.s.).

**Assumption 3.** $Z_t$ has continuous probability density function (pdf) $f(z)$.

**Assumption 4.** $f(z)$ and $m(z)$ have bounded derivatives of total order $s$.

**Assumption 5.** The functions $\omega_{ij}(z) = E(U_{it}U_{jt}|Z_t = z)$, $i, j = 1, 2, \dots$, are uniformly bounded and continuous.

Strictly, these and other assumptions need to hold only at those $z$ at which $m$ is to be estimated, but for simplicity we present them globally.

**Assumption 6.** $k(u) \in \mathcal{K}_s$, where $s \geq 2$.

**Assumption 7.** As $T \to \infty$, $a + (Ta^q)^{-1} \to 0$.

Let $f_j(z, u)$ denote the joint pdf of $(Z_t, Z_{t+j})$, $j \neq 0$, and $f_{j,k}(z, u, w)$ denote the joint pdf of $(Z_t, Z_{t+j}, Z_{t+j+k})$, $j \neq 0$, $j + k \neq 0$. Denote by $C$ a generic positive finite constant.