Available online at www.sciencedirect.com

ScienceDirect

Journal of Econometrics 144 (2008) 257–275

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

# Nonparametric estimation and testing of fixed effects panel data models

Daniel J. Henderson[a,*], Raymond J. Carroll[b], Qi Li[c,d]

[a]Department of Economics, State University of New York at Binghamton, Binghamton, NY 13902-6000, USA
[b]Department of Statistics, Texas A&M University, College Station, TX 77843-3134, USA
[c]Department of Economics, Texas A&M University, College Station, TX 77843-4228, USA
[d]Department of Economics, Tsinghua University, Beijing 100084, PR China

Received 11 May 2006; received in revised form 17 January 2008; accepted 31 January 2008
Available online 11 February 2008

**Abstract**

In this paper we consider the problem of estimating nonparametric panel data models with fixed effects. We introduce an iterative nonparametric kernel estimator. We also extend the estimation method to the case of a semiparametric partially linear fixed effects model. To determine whether a parametric, semiparametric or nonparametric model is appropriate, we propose test statistics to test between the three alternatives in practice. We further propose a test statistic for testing the null hypothesis of random effects against fixed effects in a nonparametric panel data regression model. Simulations are used to examine the finite sample performance of the proposed estimators and the test statistics.
Published by Elsevier B.V.

*Keywords:* Fixed effects models; Model specification tests; Nonparametric kernel method; Panel data; Partially linear model; Profile method; Random effects models; Semiparametric efficiency bound

## 1. Introduction

Nonparametric and semiparametric kernel methods are increasingly popular tools for statisticians/econometricians. Researchers have begun to gravitate towards nonparametric and semiparametric methods when there is little prior knowledge on specific (regression) functional forms or some known parametric specifications are deemed inadequate for the problem at hand. This often occurs when formal rejection of a parametric model yields no clues as to the direction in which to search for an improved parametric model. This growing popularity of nonparametric methods stems from their ability to relax functional form assumptions of an unknown model and let the data determine a function tailored to the data. This capacity to potentially reveal structure in the data that may be missed by common parametric specifications has encouraged growth in a variety of areas of statistics and econometrics.

The estimation of panel data models is no exception. The focus has been on both semiparametric (e.g. see Ke and Wang, 2001; Li and Stengos, 1996; Ullah and Roy, 1998) and nonparametric estimation of random

*Corresponding author. Tel.: +1 607 777 4480; fax: +1 607 777 2681.
*E-mail addresses:* djhender@binghamton.edu (D.J. Henderson), carroll@stat.tamu.edu (R.J. Carroll), qi@econmail.tamu.edu (Q. Li).

0304-4076/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.jeconom.2008.01.005

effects models (e.g. see Henderson and Ullah, 2005; Lin and Carroll, 2000, 2001, 2006; Lin et al., 2004; Lin and Ying, 2001; Ruckstuhl et al., 2000; Wang, 2003; Wu and Zhang, 2002). Estimation of these types of models is appropriate when the individual effect is independent of the regressors. This is common in many applications, where researchers often treat any unobserved individual heterogeneity as being distributed independently of the regressors. However, random effects estimators are inconsistent if the true model is one with fixed effects, i.e., individual effects which are correlated with the regressors (e.g. see Wooldridge, 2002). Indeed, economists often view the assumptions for the random effects model as being unsupported by the data. In light of this we seek to develop both nonparametric and semiparametric fixed effects estimation procedures. These procedures will be consistent under either the random or fixed effects assumptions.

We present both nonparametric and semiparametric models which either take or do not take the correlation structure into account when estimating a fixed effects nonparametric/semiparametric panel data model. Our results show that for the nonparametric model, incorporating or ignoring the within-subject correlation leads to consistent estimation results. This is shown with a sketch of the proof similar to that in Lin and Carroll (2006). However, incorporation of the correlation leads to an improvement in the estimated variance when the number of time periods is greater than two. For the semiparametric partially linear model, we also find that taking into account the correlation structure leads to efficient estimation of the finite dimensional (parametric) parameter.

Given that nonparametric estimators suffer from the curse of dimensionality, it is desirable to apply the consistent estimator with the fastest rate of convergence. Although nonparametric models are consistent under minimal assumptions, their rate of convergence is relatively slow. In contrast, semiparametric models allow for $\sqrt{n}$-convergent estimation of the parametric components, and parametric models allow all parameters to be estimated at that rate when their respective functional form restrictions are appropriate. To choose between parametric, semiparametric and nonparametric alternatives we propose in Section 4.1 tests between these three models, using a simple and practical bootstrap testing approach.

The question of whether to use random or fixed effects naturally arises with panel data. We know that when the individual effect is correlated with any of the regressors, the random effects estimator becomes biased and inconsistent. The fixed effect estimator wipes out these individual effects and leads to consistent estimates. On the other hand, if the individual effects are independent of the regressors, both estimators are consistent. In this case the random effects estimator is more efficient. This trade-off is common in econometrics and is often resolved using a testing procedure. In Section 4.2 we develop a Hausman style test for the presence of fixed versus random effects. We suggest a separate bootstrap procedure for the implementation of this test in practice.

The remainder of the paper is organized as follows: Section 2 gives the nonparametric estimation procedures when we both account for and ignore the correlation structure. Section 3 generalizes the results to the case of a semiparametric partially linear model. In Section 4 we propose test statistics for testing between parametric, semiparametric and nonparametric alternatives as well as a test statistic for testing random effects against fixed effects in nonparametric panel data regression models. Section 5 examines the finite sample properties with a small Monte Carlo study. Finally, Section 6 gives concluding remarks.

## 2. Fixed effects nonparametric panel data models

Consider the following nonparametric panel data regression model with fixed effects:

$$Y_{it} = \theta(Z_{it}) + \mu_i + v_{it} \quad (i = 1, \ldots, n; t = 1, \ldots, m), \tag{1}$$

where the functional form of $\theta(\cdot)$ is not specified. The covariate $Z_{it} = (Z_{it,1}, \ldots, Z_{it,q})$ is of dimension $q$, and all other variables are scalars. The random errors $v_{it}$ are assumed to be i.i.d. with a zero mean, finite variance and independent of $Z_{it}$ for all $i$ and $t$.[1] Further, $\mu_i$ has a zero mean and finite variance. We allow $\mu_i$ to be correlated with $Z_{it}$ with an unknown correlation structure. Hence, (1) is a fixed effects model. Alternatively, when $\mu_i$ is assumed to be uncorrelated with $Z_{it}$, model (1) is a random effects model. Note that we only consider the balanced data case in this paper for notational simplicity. The results of this paper can be generalized to the unbalanced data case.

[1] The independence between $v_{it}$ and $Z_{it}$ can be relaxed so that $v_{it}$ and $Z_{it}$ are mean independent ($\mathrm{E}(v_{it}|Z_{it}) = 0$). Simulations not reported in the paper were performed to confirm this. These results are available upon request.