Available online at www.sciencedirect.com

ScienceDirect

Journal of Econometrics 144 (2008) 193–218

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Local polynomial estimation of nonparametric simultaneous equations models

Liangjun Su[a], Aman Ullah[b,*]

[a]School of Economics, Singapore Management University, Singapore 178903, Singapore
[b]Department of Economics, University of California, Riverside, CA 92521-0427, USA

Available online 26 January 2008

Abstract

We define a new procedure for consistent estimation of nonparametric simultaneous equations models under the conditional mean independence restriction of Newey et al. [1999. Nonparametric estimation of triangular simultaneous equation models. Econometrica 67, 565–603]. It is based upon local polynomial regression and marginal integration techniques. We establish the asymptotic distribution of our estimator under weak data dependence conditions. Simulation evidence suggests that our estimator may significantly outperform the estimators of Pinkse [2000. Nonparametric two-step regression estimation when regressors and errors are dependent. Canadian Journal of Statistics 28, 289–300] and Newey and Powell [2003. Instrumental variable estimation of nonparametric models. Econometrica 71, 1565–1578].

JEL classification: C13; C14; C22

Keywords: Additive nonparametric regression; Instrumental variables; Local polynomial regression; Structural models

1. Introduction

There are many occasions in econometrics where knowledge of the structural relationship among dependent variables is required to answer questions of interest. As Newey et al. (1999) put it, structural estimation is important because we need it to account correctly for endogeneity that comes from individual choice or market equilibrium. Often, economic theory does not imply tight functional form specifications for structural models so that it is useful to consider nonparametric structural models and their estimation.

Nonparametric structural models were first considered in Roehrig (1988), and Newey and Powell (1989), among others. Assuming that the errors are independent of the instruments, Roehrig (1988) gives identification results for a system of equations. Under the weaker condition that the disturbance has conditional mean zero given the instruments, Newey and Powell (1989) consider both identification and estimation problems. Following this latter paper, Brown and Matzkin (1998), Newey et al. (1999), Pinkse (2000), Darolles et al. (2000), Horowitz (2005), and Imbens and Newey (2006) consider identification and

*Corresponding author. Tel.: +1 909 827 1591.
E-mail addresses: lsu@gsm.pku.edu.cn (L. Su), aman.ullah@ucr.edu (A. Ullah).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2008.01.002

estimation of different nonparametric models under various restrictions. For example, Pinkse (2000) considers estimation of a structural model by assuming the independence between the instrumental variable and the error terms in both the structural model and reduced model.

In this paper, we consider the regression model of Newey et al. (1999):

$$\begin{cases} Y = g(X, Z_1) + \varepsilon, \quad Z = (Z_1', Z_2')', \\ X = h(Z) + U, \quad \mathrm{E}(U|Z) = 0, \quad \mathrm{E}[\varepsilon|Z, U] = \mathrm{E}[\varepsilon|U], \end{cases} \tag{1.1}$$

where Y is an observable scalar random variable, g denotes the true, unknown structural function of interest, X is $d_x \times 1$ vector of explanatory variables, Z_1 and Z_2 are $d_1 \times 1$ and $d_2 \times 1$ vectors of instrumental variables, $h \equiv (h_1, \ldots, h_{d_x})'$ is a $d_x \times 1$ vector of functions of the instruments Z, and U and ε are disturbances. We are interested in estimating g and its derivatives consistently.

Newey et al. (1999) show that g is identified up to an additive constant if there is no functional relationship between (X, Z_1) and U.[1] They employ series approximations that exploit the additive structure of the model and propose a two-stage estimator of g. They derive consistency and asymptotic normality results for functionals of their estimator. By contrast, Newey and Powell (2003) study the estimation of g in (1.1) under the restrictions that $\mathrm{E}[\varepsilon|Z] = 0$ and $\mathrm{E}(U|Z) = 0$, and give identification results. Based on sieve approximations, they propose an estimator of g that is a nonparametric analog to the familiar two-stage least squares (2SLS) estimator for linear models with endogenous regressors and prove a consistency result for their estimator. Nevertheless, neither the consistency rate nor the normality of the proposed estimator is obtained.

In this paper, we propose a local polynomial procedure for estimating $g(\cdot)$ in (1.1) that is based on the following observation:

$$\begin{aligned} \mathrm{E}[Y|X, Z, U] &= g(X, Z_1) + \mathrm{E}[\varepsilon|X, Z, U] \\ &= g(X, Z_1) + \mathrm{E}[\varepsilon|X - h(Z), Z, U] \\ &= g(X, Z_1) + \mathrm{E}[\varepsilon|Z, U] \\ &= g(X, Z_1) + \mathrm{E}[\varepsilon|U]. \end{aligned} \tag{1.2}$$

Thus it follows from the law of iterated expectations that

$$m(X, Z_1, U) \equiv \mathrm{E}[Y|X, Z_1, U] = g(X, Z_1) + \mathrm{E}[\varepsilon|U]. \tag{1.3}$$

Like Newey et al. (1999), our procedure can estimate $g(\cdot)$ consistently up to an additive constant that explores the additive structure in the above model. If the realizations of U were observable, the model is simply the additive model widely studied in the literature. One can adopt the marginal integration technique (e.g., Linton and Härdle, 1996) to estimate g. Further, Linton (1997, 2000) defines a two-step estimator for generalized additive nonparametric regression models that is more efficient than the marginal integration estimator. Thus one can go one step further to obtain a more efficient estimator of g. Here, because the realizations of U are not observed, we replace them by the residuals obtained by regressing X on Z nonparametrically. We show that such a replacement does not affect the first-order asymptotic property of the resulting estimator.[2]

Like Pinkse (2000) we will allow for weak data dependence in our estimation procedure. A typical application is the estimation of the Kuznets curve using time series data, which relates economic growth to economic inequality. Lundberg and Squire (2003) emphasize the simultaneous evolution of growth and inequality. Farrell et al. (1999) model the structural relationship between lottery sales and the expected value using time series data. For more studies on the structural time series models, see Zellner and Palm (2004). Given the unknown nonlinear relationship between the variables of interest in all these works, one may like to model them nonparametrically.

The rest of the paper is organized as follows. We introduce our estimator and its asymptotic distribution theory in Section 2. We report some Monte Carlo simulation results in Section 3. Section 4 concludes. All proofs are given in the Appendix.

[1]The conditional moment restrictions in (1.1) are much weaker than those imposed in Pinkse (2000).

[2]In a different but relevant context, Li and Wooldridge (2002) show that $\sqrt{n}$-consistent estimation results of the finite dimensional parameter in a partially linear model can be generalized to the case of generated regressors with weakly dependent data.