A Nonparametric Random Effects Estimator

Daniel J. Henderson and Aman Ullah*
Department of Economics
University of California, Riverside

May 1, 2003

Abstract

This paper considers the problem of improving the estimation of a one-way random effects error component model. A nonparametric estimator is proposed, its structure is defined and its asymptotic properties are proven. Monte Carlo shows that the proposed estimator performs almost as well as the parametric estimator in linear technology, but drastically outperforms the parametric estimator when the technology becomes nonlinear.

Keywords: Generalized Least Squares, Nonparametric, Random Effects
JEL Classification: C1, C4

*Acknowledgments: The authors would like to thank Badi Baltagi and Jang-Ting Guo for useful comments. Contact Information: Daniel J. Henderson, Department of Economics, University of California, Riverside, CA 92521, (909) 787-5037, Fax: (909) 787-5685, e-mail: djhenderson8@hotmail.com. Aman Ullah, Department of Economics, University of California, Riverside, CA 92521, (909) 827-1591, Fax (909) 787-5685, e-mail: ullah@mail.ucr.edu.

1 Introduction

Economic research has been enriched by the increased availability of panel (longitudinal) data that measure cross-sectional units over a period of time.[1] The primary advantage with panel over cross-sectional data is that the researcher has increased flexibility when modeling differences in the cross-sectional units. The basic framework for this analysis is the following model:

$$y_{it} = \alpha + x_{it}\beta + u_i + v_{it},$$

where $i = 1, 2, ..., N$, $t = 1, 2, ..., T$, y_{it} is the endogenous variable, α is the constant term, x_{it} is a matrix of k exogenous variables, β is a $k \times 1$ unknown parameter, u_i is known as the individual effect and v_{it} is the random error. This is known as the one-way error component (correction) model.[2] The individual effect is what separates the one-way error component model from the classical linear regression model, it is constant over time and is specific to each cross-sectional unit i.

There are three basic frameworks used to generalize and estimate this model: random effects, fixed effects and maximum likelihood estimation. The random effects approach treats u_i as a group specific disturbance.. This framework is most appropriate when the cross-sectional units are believed to be sampled from a large population. The fixed effects approach treats the individual effect as a group specific constant term within the regression model. This approach is most appropriate when the cross-sectional units are the complete set of the population, meaning the researcher can be confident that the differences between the cross-sectional units can be viewed as parametric shifts of the regression function. Maximum likelihood estimation treats the u_i as random disturbances that follow a particular distribution. This approach is most appropriate when the distribution of the individual effect is known. Since most economic data is a sample taken from a larger population, in this paper we consider the estimation of random effects models where u_i is random.[3]

[1] For the benefits and limitations of standard panel data and error correction models as well as econometric estimation, one should consult Baltagi (2001) or Hsiao (2002).

[2] The term, one-way error component model, comes from the structure of the error terms $\varepsilon_{it} = u_i + v_{it}$, as opposed to the two-way error component model in which a parameter only indexed by time, λ_t, is added. Only the one-way error component model will be discussed in this paper and the two-way model is left for future research.

[3] For further discussion and tests to detmine which method is appropiate as well as estimation by the other techniques, see Greene (2002).

In the case of linear regressions, like the one above, a particular concern has been with the linearity of the functional form connecting the variables of the model. Often the true technology is unknown and linear regressions are performed without economic reasoning due to their straightforward estimation procedures and well-known properties. This concern initially spawned an interest in transformations of the endogenous and exogenous variables, leading to the use of flexible specifications, such as the translog functional form. Although approaches such as these have served econometrics well, there has always been some worry that the functional form might be more complex. Thus, it is worthwhile considering nonparametric estimation if the functional form is unknown. The basic idea behind nonparametric estimation is to approximate the technology arbitrarily close. Unfortunately, the literature, up to this point does not possess a nonparametric kernel estimator for the one-way error component random effects model. This paper presents such an estimator, one in which no functional form is associated with any of the regressors.[4]

This paper is organized as follows: Section 2 gives the model, notation, proposes a new estimator and derives the theoretical estimates. Section 3 provides the Monte Carlo setup and summarizes the results of the experiments. Finally, Section 4 concludes the paper.

2 The Model

Let us consider a nonparametric one-way error component model as

$$y_{it} = m(x_{it}) + \varepsilon_{it}, \tag{1}$$

where $i = 1, 2, ..., N$, $t = 1, 2, ..., T$, y_{it} is the endogenous variable, x_{it} is a vector of k exogenous variables and $m(\cdot)$ is an unknown smooth function. Further, ε_{it} follows the one-way error component specification

$$\varepsilon_{it} = u_i + v_{it}, \tag{2}$$

where u_i is $i.i.d.$ $(0, \sigma_u^2)$, v_{it} is $i.i.d.$ $(0, \sigma_v^2)$ and u_i and v_{it} are uncorrelated for all q and ls, where q is different from l; $q, l \in i$ and $s \in t$.

[4] For an example of semiparametric methods, (making use of the Robinson (1988) technique and examining the parametric parameters) see Berg, Li and Ullah (2000).

Let $\varepsilon_i = [\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT}]'$ be a $T \times 1$ vector. Then $V \equiv E(\varepsilon_i \varepsilon_i')$, takes the form

$$V = \sigma_v^2 \mathbf{I}_T + \sigma_u^2 \mathbf{i}_T \mathbf{i}_T', \tag{3}$$

where $\mathbf{I}$ is an identity matrix of dimension T and $\mathbf{i}$ is a $T \times 1$ column vector of ones. Since the observations are independent over q and l, the covariance matrix for the full $NT \times 1$ disturbance vector ε, $\Omega = E(\varepsilon\varepsilon')$ is

$$\Omega = V \otimes \mathbf{I}_N. \tag{4}$$

We are interested in estimating the unknown function $m(x)$ at a point x and the slope of $m(x)$, $\beta(x) = \nabla m(x)$, where ∇ is the gradient vector of $m(x)$. The parameter $\beta(x)$ is interpreted as a varying coefficient. We consider the usual panel data situation of large N and small T.

Nonparametric kernel estimation of $m(x)$ and $\beta(x)$ can be obtained by using local linear least squares (LLLS) estimation. This is obtained by minimizing the local least squares or weighted least squares of errors

$$\sum_i \sum_t (y_{it} - X_{it}\delta(x))^2 K\left(\frac{x_{it} - x}{h}\right) = (y - X\delta(x))K(x)(y - X\delta(x)) \tag{5}$$

with respect to $m(x)$ and $\beta(x)$, where y is a $NT \times 1$ vector, X is a $NT \times (k+1)$ matrix generated by $X_{it} = (1 \quad (x_{it} - x))$, $\delta(x) = (m(x), \beta(x))'$ is a $(k+1) \times 1$ vector, $K(x)$ is an $NT \times NT$ diagonal matrix of kernel (weight) functions $K(\frac{x_{it}-x}{h})$ and h is the bandwidth (smoothing) parameter. Generally kernel functions can be any probability function having a finite second moment (here we use the standard normal kernel). The estimator so obtained is

$$\widehat{\delta}(x) = (X'K(x)X)^{-1}X'K(x)y \tag{6}$$

The estimator of $m(x)$ is then given by $\widehat{m}(x) = (1 \quad 0)\widehat{\delta}(x)$, whereas $\widehat{\beta}(x)$ can be extracted from $\widehat{\delta}(x)$ as $\widehat{\beta}(x) = (0 \quad 1)\widehat{\delta}(x)$. The estimator in (6) is called LLLS estimator. Asymptotic normality for the cross-sectional case is proven by Li and

Woolridge (2000) and Kniesner and Li (2002) derive a proof for the case of panel data.[5]

The LLLS estimator in (6) however ignores the information contained in the disturbance vector covariance matrix Ω. In view of this we introduce a new estimator, local linear generalized least squares (LLGLS) estimator, by minimizing

$$(y - X\delta(x))'\sqrt{K(x)}\Omega^{-1}\sqrt{K(x)}(y - X\delta(x)) \tag{7}$$

with respect to $\delta(x)$. This gives the estimator $d(x)$

$$d(x) = (X'\sqrt{K(x)}\Omega^{-1}\sqrt{K(x)}X)^{-1}X'\sqrt{K(x)}\Omega^{-1}\sqrt{K(x)}y. \tag{8}$$

The objective function in (7) amounts to doing GLS (linear) fits to the points local to x.[6] The LLGLS estimator in (8) however depends upon the unknown parameters σ_u^2 and σ_v^2. An estimator of σ_v^2 is obtained by using the within estimator. The exact form of this $\widehat{\sigma}_v^2$ is

$$\widehat{\sigma}_v^2 = \frac{1}{NT}\sum_i\sum_t (\widehat{v_{it} - \overline{v}_i})^2 \tag{9}$$

where

$$\sum_i\sum_t (\widehat{v_{it} - \overline{v}_i})^2 = \sum_i\sum_t [(y_{it} - \overline{y}_i) - (x_{it} - \overline{x}_i)\beta^*(x_{it})]^2 \tag{10}$$

in which $\overline{y}_i = \frac{1}{T}\sum_t y_{it}$, $\overline{x}_i = \frac{1}{T}\sum_t x_{it}$ and

$$\beta^*(x) = [\sum_i\sum_t (x_{it} - \overline{x}_i)'(x_{it} - \overline{x}_i)K_{it}]^{-1}\sum_i\sum_t (x_{it} - \overline{x}_i)'(y_{it} - \overline{y}_i)K_{it}, \tag{11}$$

[5] For more information on the choices of K and h see Fan and Gijbels (1992) and Pagan and Ullah (1999).

[6] We also consider an alternative nonparametric estimator which takes the form: $\widehat{\delta}(x)_{ANPFGLS} = (X'\widehat{\Omega}^{-\frac{1}{2}}K(X)\widehat{\Omega}^{-\frac{1}{2}}X)^{-1}(X'\widehat{\Omega}^{-\frac{1}{2}}K(X)\widehat{\Omega}^{-\frac{1}{2}}Y)$. Although it looks similar, there is an inherent flaw in the way the Alternative Nonparametric Feasible Generalized Least Squares (ANPFGLS) estimator transforms the data. By simply looking at the matrix structure it becomes evident. It shows that in the ANPFGLS setup the variables are first adjusted for heteroskedasticity (i.e. $X'\widehat{\Omega}^{-\frac{1}{2}} = X^*$), and then run through the kernel (i.e. $X^{*\prime}K(X)$). The basic idea behind the nonparametric estimator is to provide a separate estimate for each value of x, the formulation of the ANPFGLS gives a separate estimate for each value of x^*. Therefore it is this process of the variables being adjusted for heteroskedasticy before they are smoothed which causes the mispecification, whereas the NPFGLS estimators first transform the data correctly (i.e. $X'K(X)$) and then proceed to adjust it for heteroskedasticity.

where $K_{it} = K\left(\frac{x_{it}-x}{h}\right)$.

One estimate of σ_u^2 is obtained as a combination of $\widehat{\sigma}_v^2$ and $\widehat{\sigma}_\varepsilon^2$ being

$$\widehat{\sigma}_u^2 = \widehat{\sigma}_\varepsilon^2 + \frac{1}{T}\widehat{\sigma}_v^2, \tag{12}$$

in which

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{N-k}\sum_i \widehat{\overline{\varepsilon}}_i^2, \tag{13}$$

where

$$\sum_i \widehat{\overline{\varepsilon}}_i^2 = \frac{1}{N}\sum_i \left[(\overline{y}_i - \overline{y}) - (\overline{x}_i - \overline{x})\,\widetilde{\beta}(x_i)\right]^2, \tag{14}$$

where

$$\widetilde{\beta}(x) = \frac{\sum_i (\overline{y}_i - \overline{y})(\overline{x}_i - \overline{x})\,K\left(\frac{x_i - x}{h}\right)}{\sum_i (\overline{x}_i - \overline{x})^2\,K\left(\frac{x_i - x}{h}\right)}. \tag{15}$$

where $\overline{y} = \frac{1}{N}\sum_i \overline{y}_i$ and $\overline{x} = \frac{1}{N}\sum_i \overline{x}_i$.

Alternative estimators of σ_u^2 and σ_v^2 can be obtained by noting that $V(\varepsilon_{it}) \equiv \sigma_\varepsilon^2 = \sigma_u^2 + \sigma_v^2$ and

$$cov(\varepsilon_{qt}, \varepsilon_{lt'}) = \sigma_u^2 \text{ for } q \neq l$$

and zero otherwise. Thus

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{NT}\sum_i\sum_t \widehat{\varepsilon}_{it}^2 \tag{16}$$

and

$$\widehat{\sigma}_u^2 = \frac{1}{N(T-1)}\sum_i\sum_t\sum_{t\neq t'} \widehat{\varepsilon}_{it}\widehat{\varepsilon}'_{it'} \tag{17}$$

where $\widehat{\varepsilon}_{it} = y_{it} - X_{it}\widehat{\delta}(x_{it})$ is the LLLS residual based on the first stage estimator of $\widehat{\delta}(x)$ in (6).[7]

[7] Note that estimation of the third variance term is straightforward. Also, it has been noted by Maddala and Mount (1973) and Taylor (1980) that more efficient esimation of the variance components does not necessilarly lead to more effecient estimates of $m(x)$ and $\beta(x)$.

Substituting the estimators of σ_u^2 and σ_v^2 from (9) and (12) or (16) and (17) into (8) gives a feasible linear generalized least squares (FLGLS) or nonparametric feasible generalized least squares (NPFGLS) estimator as

$$\widehat{\delta}(x)_{NPFGLS} = (X'\sqrt{K(x)}\widehat{\Omega}^{-1}\sqrt{K(x)}X)^{-1}X\sqrt{K(x)}\widehat{\Omega}^{-1}\sqrt{K(x)}y.$$

Consistency of $\widehat{\delta}(x)$ is straightforward under the standard regularity conditions, but asymptotic normality requires somewhat stronger assumptions as stated in Li and Woolridge (2000, p. 340). Asymptotic normality is established under the following theorem.

THEOREM:

Under the standard assumptions

$$D(NT)\left(\widehat{\delta}(x) - \delta(x) - \begin{pmatrix} h^2\mu_k W \\ 0 \end{pmatrix}\right) \to N(0, \Sigma_x)$$

where $D(NT) = \begin{pmatrix} (NTh^k)^{\frac{1}{2}} & 0 \\ 0 & (NTh^{k+2})^{\frac{1}{2}} \end{pmatrix}$, $\mu_k = \frac{1}{2}c_k tr(m''(x))$, $W = tr(\Omega^{-1})$, $\Sigma_x = \begin{pmatrix} d_k\sigma_\varepsilon^2/f(x)W & 0 \\ 0 & v_k\sigma_\varepsilon^2 I_k/c_k^2 f(x)W \end{pmatrix}$, $c_k = \int K(\psi)\psi\psi' d\psi$, $d_k = \int K^2(\psi)d\psi$ and $v_k = \int K^2(\psi)\psi\psi' d\psi$, where ψ is defined as $\frac{x_{it}-x}{h}$. The proof is given in the appendix. It loosely follows the method used by Li and Wooldridge (2000) for the cross-sectional case. Our proof differs by incorporating information about the variance parameters. Also note that the proof is given for the most general case (LLGLS) and is easily modified to satisfy the above scenario.

3 Monte Carlo Results

Although asymptotic results give clues as to the performance of the estimators, most economic panel data is finite. This section uses Monte Carlo simulations to examine the finite sample performance of the proposed NPFGLS estimator. Following the methodology of Baltagi, Chang and Li (1992), the following data generating process is used:

$$y_{it} = \alpha + x_{it}\beta + x_{it}^2\gamma + u_i + v_{it},$$

where x_{it} is generated by the method of Nerlove (1971).[8] The value of α is chosen to be 5, β is chosen to be 0.5 and γ takes the values of 0 (linear technology) and 2 (quadratic technology). The distribution of u_i and v_{it} are generated separately as i.i.d. Normal. Total variance of $\sigma_v^2 + \sigma_u^2 = 20$ and $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_v^2)$ is varied to be 0.1, 0.4 and 0.8.

For comparison, we compute the following estimators of δ:

(I) Parametric (linear) Feasible GLS (FGLS) estimator

$$\widehat{\delta}_{FGLS} = (X'\widehat{\Omega}^{-1}X)^{-1}(X'\widehat{\Omega}^{-1}Y).$$

(II) NPFGLS estimator

$$\widehat{\delta}_{NPFGLS} = (X'\sqrt{K(x)}\widehat{\Omega}^{-1}\sqrt{K(x)}X)^{-1}X\sqrt{K(x)}\widehat{\Omega}^{-1}\sqrt{K(x)}y.$$

Reported are the estimated bias and mean squared error (MSE) for each estimator. These are computed via $Bias(\widehat{m}) = M^{-1}\sum\limits_j(\widehat{m}_j - m^*)$, and $MSE(\widehat{m}) = M^{-1}\sum\limits_j (\widehat{m}_j - m^*)^2$ where M is the number of replications, $\widehat{m}_j$ is the estimated value of m^* at the jth replication and $m^* = \alpha + \overline{x}\beta + \overline{x}^2\gamma$. Similarly for the varying coefficient parameter, $Bias(\widehat{\beta}) = M^{-1}\sum\limits_j(\widehat{\beta}_j - \beta^*)$ and $MSE(\widehat{\beta}) = M^{-1}\sum\limits_j(\widehat{\beta}_j - \beta^*)^2$, where $\widehat{\beta}_j$ is the estimated value of β^* at the jth replication and $\beta^* = \beta + 2\overline{x}\gamma$. $M = 1000$ is used in all simulations, T is varied to be 3 and 5, while N takes the values 10, 20 and 50. The simulation results are given in Tables 1 through 4. The smallest MSE for each case (for a given N, T, ρ and γ) is shown as a boldface number.[9]

Tables 1 and 2 report the result for $\gamma = 0$ (linear technology). In each case, the parametric estimators outperform the nonparametric estimators in MSE. This result is expected since the true underlying technology is linear and because nonparametric

[8] The x_{it} were generated as follows: $x_{it} = 0.1t + 0.5x_{it-1} + w_{it}$, where $x_{i0} = 10 + 5w_{i0}$ and $w_{it} \sim U[-\frac{1}{2}, \frac{1}{2}]$.

[9] In the tables only local estimates of $\widehat{\delta}$ are given. The specific form being

$$\widehat{\delta}(\overline{x})_{NPFGLS} = (X'K(\overline{x})^{\frac{1}{2}}\widehat{\Omega}^{-1}K(\overline{x})^{\frac{1}{2}}X)^{-1}XK(\overline{x})^{\frac{1}{2}}\widehat{\Omega}^{-1}K(\overline{x})^{\frac{1}{2}}y,$$

where $K(\overline{x}) = I_{NT} \times K\left(\frac{x_{it}-\overline{x}}{h}\right)$. These estimates are similar to the global ones, except that rather than being evaluated at each value of x and then averaged, they are only evaluated at the mean value of x. Being that these estimates are evaluated at the mean of x, they appear to perform better in Monte Carlo exercises that evaluate at the mean of x. It should be noted however that this method rarely outperforms the global measures in empirical data and is usually only used for computational ease. Further, replacing these with the global estimates does not affect the conclusions of the paper.

estimators are known to have small sample bias. Contrary to the first two tables, the results of Tables 3 and 4 are in support of the NPFGLS estimators. The NPFGLS estimators outperform the parametric estimators drastically in both $Bias$ and MSE. This table shows the major drawback of the parametric type estimators and the strengths of the nonparametric type estimators. When the technology is complex, parametric estimators are often mispecified and the nonparametric estimators, with their complete flexibility, are better able to adapt.

4 Concluding Remarks

This paper examines the problem of improving the estimation of a one-way random effects error component model.[10] A nonparametric estimator is proposed, its structure is defined, asymptotic properties proven and its finite sample results are generated through a Monte Carlo exercise. The Monte Carlo results of section 3 show that the parametric estimators provide slightly smaller MSE when the true underlying technology is linear and correctly specified. Although less efficient in the aforementioned exercises, the NPFGLS estimators provide acceptable estimation. On the other hand, when the technology becomes complex, the NPFGLS estimators perform best, with the MSE of the linear parametric estimator up to 70 times that of the NPFGLS estimator. Thus, it is suggested that the NPFGLS estimators be used in practice because the true underlying technology is usually unknown; and as Tables 3 and 4 demonstrate, the consequences of choosing the wrong estimator may be quite severe.

[10] Throughout the paper, the existence of random individual effects is assumed. In practice one may want to test for the existence of random individual effects.

Appendix

Following Li and Woolridge (2000) we rewrite $d(x)$ as

$$d(x) = \begin{pmatrix} \sum_i K_i w_{ii} G_N \binom{1}{x_i - x} \begin{pmatrix} 1 & (x_i-x)' \end{pmatrix} \\ + \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} G_N \binom{1}{x_i - x} \begin{pmatrix} 1 & (x_i-x)' \end{pmatrix} \end{pmatrix}^{-1}$$

$$\left(\sum_i K_i w_{ii} G_N \begin{pmatrix} 1 \\ x_i - x \end{pmatrix} y_i + \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} G_N \begin{pmatrix} 1 \\ x_i - x \end{pmatrix} y_i \right)$$

where

$$\begin{aligned} G_N &= \begin{pmatrix} h^2 & 0 \\ 0 & I_k \end{pmatrix}, \\ I_k &= G_N^{-1} G_N, \\ K_i &= K\left(\frac{x_i - x}{h} \right) \end{aligned}$$

and

$$\Omega_{ij}^{-1} \equiv w_{ij}.$$

Multiplying both the numerator and denominator by $\frac{1}{Nh^{k+2}}$ and substituting the Taylor expansion $\left(y_i = \begin{pmatrix} 1 & (x_i - x)' \end{pmatrix} \delta(x) + (x_i - x)\, m''(x)\,(x_i - x)'/2 + R_m(x_i, x) + \varepsilon_i\right)$ into the above equation gives

$$d(x) = \begin{pmatrix} \frac{1}{Nh^{k+2}} \sum_i K_i w_{ii} \binom{h^2}{x_i - x} \begin{pmatrix} 1 & (x_i-x)' \end{pmatrix} \\ + \frac{1}{Nh^{k+2}} \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} \binom{h^2}{x_i - x} \begin{pmatrix} 1 & (x_i-x)' \end{pmatrix} \end{pmatrix}^{-1}$$

$$\begin{pmatrix} \frac{1}{Nh^{k+2}} \sum_i K_i w_{ii} \binom{h}{x_i - x} \begin{pmatrix} \begin{pmatrix} 1 & (x_i-x)' \end{pmatrix} \delta(x) \\ +(x_i-x)' m''(x)(x_i-x)/2 + R_m(x_i,x) + \varepsilon_i \end{pmatrix} \\ + \frac{1}{Nh^{k+2}} \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} \binom{h^2}{x_i - x} \begin{pmatrix} \begin{pmatrix} 1 & (x_i-x)' \end{pmatrix} \delta(x) \\ +(x_i-x)' m''(x)(x_i-x)/2 + R_m(x_i,x) + \varepsilon_i \end{pmatrix} \end{pmatrix}.$$

After simplifying this expression it becomes obvious that

$$
\begin{aligned}
d(x) &= \delta(x) + \left(\begin{array}{c} \frac{1}{Nh^{k+2}} \sum_i K_i w_{ii} \begin{pmatrix} h^2 & h^2(x_i - x)' \\ (x_i - x) & (x_i - x)(x_i - x)' \end{pmatrix} \\ + \frac{1}{Nh^{k+2}} \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} \begin{pmatrix} h^2 & h^2(x_i - x)' \\ (x_i - x) & (x_i - x)(x_i - x)' \end{pmatrix} \end{array} \right)^{-1} \\
&\quad \left(\begin{array}{c} \frac{1}{Nh^{k+2}} \sum_i K_i w_{ii} \binom{h^2}{x_i - x} \left((x_i - x)' m''(x)(x_i - x)/2 + R_m(x_i, x) + \varepsilon_i \right) \\ + \frac{1}{Nh^{k+2}} \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} \binom{h^2}{x_i - x} \left((x_i - x)' m''(x)(x_i - x)/2 + R_m(x_i, x) + \varepsilon_i \right) \end{array} \right) \\
&\equiv \delta(x) + \left(A^{1,x} \right)^{-1} \left(A^{2,x} + A^{3,x} \right) + (s.o.)
\end{aligned}
$$

where

$$
A^{1,x} = \begin{array}{c} \frac{1}{Nh^{k+2}} \sum_i K_i w_{ii} \begin{pmatrix} h^2 & h^2(x_i - x)' \\ (x_i - x) & (x_i - x)(x_i - x)' \end{pmatrix} \\ + \frac{1}{Nh^{k+2}} \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} \begin{pmatrix} h^2 & h^2(x_i - x)' \\ (x_i - x) & (x_i - x)(x_i - x)' \end{pmatrix} \end{array},
$$

$$
\begin{aligned}
A^{2,x} &= \frac{1}{Nh^{k+2}} \sum_i K_i w_{ii} \binom{h^2}{x_i - x} (x_i - x)' m''(x)(x_i - x)/2 \\
&\quad + \frac{1}{Nh^{k+2}} \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} \binom{h^2}{x_i - x} (x_i - x)' m''(x)(x_i - x)/2,
\end{aligned}
$$

$$
A^{3,x} = \frac{1}{Nh^{k+2}} \sum_i K_i w_{ii} \binom{h^2}{x_i - x} \varepsilon_i + \frac{1}{Nh^{k+2}} \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} \binom{h^2}{x_i - x} \varepsilon_i
$$

and

$$
s.o. = \frac{1}{Nh^{k+2}} \sum_i K_i w_{ii} \binom{h^2}{x_i - x} R_m(x_i, x) + \frac{1}{Nh^{k+2}} \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} \binom{h^2}{x_i - x} R_m(x_i, x)
$$

where $(s.o.)$ has smaller order than $\left(A^{1,x}\right)^{-1}\left(A^{2,x}\right)$. We can now rewrite the expression as

$$
D(N)(d(x) - \delta(x)) = D(N) \left(A^{1,x} \right)^{-1} \left(A^{2,x} + A^{3,x} \right) + (s.o.).
$$

We will prove the theorem if we prove the following four statements:

(i) $D(N) \left(A^{1,x} \right)^{-1} \left(A^{2,x} + A^{3,x} \right) = D(N) M^{-1} (A^{2,x} + A^{3,x}) + o_p(1)$

(ii) $D(N) M^{-1} (A^{2,x} + A^{3,x}) = RD(N)(A^{2,x} + A^{3,x}) + o_p(1)$

(iii) $D(N)A^{2,x} = \binom{(Nh^{k+4})^{\frac{1}{2}}\mu_k f(x)W}{0} + o_p(1)$

(iv) $D(N)A^{3,x} \to N(0,V)$ in dist

where

$$M = \begin{pmatrix} f(x)W & 0 \\ c_k f'(x)W & c_k f(x)WI_k \end{pmatrix}.$$

Further

$$R = diag(M^{-1})$$

and

$$V = \begin{pmatrix} d_k\sigma_\varepsilon^2 f(x)W & 0 \\ 0 & v_k\sigma_\varepsilon^2 f(x)WI_k \end{pmatrix}.$$

Proof of (i). Note that $D(N)\left(A^{1,x}\right)^{-1}(A^{2,x}+A^{3,x}) = D(N)M^{-1}(A^{2,x}+A^{3,x}) + D(N)\left(\left(A^{1,x}\right)^{-1} - M^{-1}\right)(A^{2,x}+A^{3,x})$. Thus, we only need to show that $D(N)\left(\left(A^{1,x}\right)^{-1} - M^{-1}\right)(A^{2,x}+A^{3,x}) = o_p(1)$. This can be shown by first examining the asymptotic behavior of each element in $A^{1,x}$. Defining

$$A^{1,x} = \begin{pmatrix} A_{11}^{1,x} & A_{12}^{1,x} \\ A_{21}^{1,x} & A_{22}^{1,x} \end{pmatrix}$$

yields

$$A_{11}^{1,x} = \frac{1}{Nh^{k+2}}\sum_i K_i w_{ii} h^2 + \frac{1}{Nh^{k+2}}\sum_{i\neq j}\sqrt{K_i}\sqrt{K_j}w_{ij}h^2,$$

$$A_{21}^{1,x} = \frac{1}{Nh^{k+2}}\sum_i K_i w_{ii}(x_i - x) + \frac{1}{Nh^{k+2}}\sum_{i\neq j}\sqrt{K_i}\sqrt{K_j}w_{ij}(x_i - x),$$

$$A_{12}^{1,x} = \frac{1}{Nh^{k+2}}\sum_i K_i w_{ii} h^2 (x_i - x)' + \frac{1}{Nh^{k+2}}\sum_{i\neq j}\sqrt{K_i}\sqrt{K_j}w_{ij}h^2(x_i - x)',$$

and

$$A_{22}^{1,x} = \frac{1}{Nh^{k+2}}\sum_i K_i w_{ii}(x_i - x)(x_i - x)' + \frac{1}{Nh^{k+2}}\sum_{i\neq j}\sqrt{K_i}\sqrt{K_j}w_{ij}(x_i - x)(x_i - x)'.$$

By taking the first element we achieve

$$
\begin{aligned}
E(A_{11}^{1,x}) &= \frac{1}{Nh^k}\sum_i E(K_i)w_{ii} + \frac{1}{Nh^k}\sum_{i\neq j} E\left(\sqrt{K_i}\sqrt{K_j}\right)w_{ij} \\
&= \frac{1}{h^k}\int K\left(\frac{x_i-x}{h}\right)f(x_i)dx_i\frac{1}{N}\sum_i w_{ii} \\
&\quad +\frac{1}{h^k}\int K\left(\frac{x_i-x}{h}\right)^{\frac{1}{2}}f(x_i)dx_i\int K\left(\frac{x_j-x}{h}\right)^{\frac{1}{2}}f(x_j)dx_j\frac{1}{N}\sum_{i\neq j} w_{ij} \\
&\to f(x)\frac{1}{N}\sum_i w_{ii} + o_p(1) \\
&= f(x)tr\left(\Omega^{-1}\right) + o_p(1) \\
&= f(x)W + o_p(1).
\end{aligned}
$$

$A_{21}^{1,x}$ is decomposed similarly as

$$
\begin{aligned}
E\left(A_{21}^{1,x}\right) &= \frac{1}{h^2}\int f(x+\psi h)K(\psi)h\psi d\psi\frac{1}{N}\sum_i w_{ii} \\
&\quad +\frac{1}{h^2}\int K\left(\frac{x_i-x}{h}\right)^{\frac{1}{2}}f(x_i)dx_i\int K\left(\frac{x_j-x}{h}\right)^{\frac{1}{2}}f(x_j)dx_j\frac{1}{N}\sum_{i\neq j} w_{ij} \\
&\to c_k f'(x)tr\left(\Omega^{-1}\right) + o_p(1) \\
&= c_k f'(x)W + o_p(1).
\end{aligned}
$$

Next, the term $A_{12}^{1,x} = h^2(A_{21}^{1,x})' = O_p(h^2)$. Finally, $A_{22}^{1,x}$ can be shown as

$$
\begin{aligned}
E\left(A_{22}^{1,x}\right) &= f(x)\int K(\psi)\psi\psi' d\psi\frac{1}{N}\sum_i w_{ii} \\
&\quad +\frac{1}{h^2}\int K\left(\frac{x_i-x}{h}\right)^{\frac{1}{2}}f(x_i)dx_i\int K\left(\frac{x_j-x}{h}\right)^{\frac{1}{2}}f(x_j)dx_j\frac{1}{N}\sum_{i\neq j} w_{ij} \\
&\to c_k f(x)tr\left(\Omega^{-1}\right)I_k + o_p(1) \\
&= c_k f(x)WI_k + o_p(1).
\end{aligned}
$$

Thus we have

$$
A^{1,x} = \begin{pmatrix} f(x)W & 0 \\ c_k f'(x)W & c_k f(x)WI_k \end{pmatrix} + o_p(1).
$$

By inverting this matrix (through the method of the partitioned inverse) we achieve

$$\left(A^{1,x}\right)^{-1} = \begin{pmatrix} \frac{1}{f(x)W} + o_p(1) & O_p(h^2) \\ \frac{-f'(x)}{f^2(x)W} + o_p(1) & \frac{I_k}{c_k f(x)W} + o_p(1) \end{pmatrix}$$

and thus

$$\left(A^{1,x}\right)^{-1} - M^{-1} = \begin{pmatrix} o_p(1) & O_p(h^2) \\ o_p(1) & o_p(1)I_k \end{pmatrix}$$

which completes the proof of (i).

Proof of (ii). This holds because the off diagonal elements of M^{-1} are $o_p(1)$. Specifically,

$$\begin{aligned} \left(Nh^{k+2}\right)^{\frac{1}{2}} A_1^{2,x} &= \left(Nh^{k+2}\right)^{\frac{1}{2}} O_p(h^2) \\ &= O_p\left(\left(Nh^{k+6}\right)^{\frac{1}{2}}\right) \\ &= o_p(1) \end{aligned}$$

and

$$\begin{aligned} \left(Nh^{k+2}\right)^{\frac{1}{2}} A_2^{2,x} &= \left(Nh^{k+2}\right)^{\frac{1}{2}} O_p\left(\left(Nh^{k}\right)^{\frac{1}{2}}\right) \\ &= O_p\left(h\right) \\ &= o_p(1). \end{aligned}$$

Proof of (iii). Premultiplying $A^{2,x}$ by $D(N)$ gives

$$\begin{aligned} D(N) &= \begin{pmatrix} \left(Nh^{k}\right)^{\frac{1}{2}} A_1^{2,x} \\ (Nh^{k+2})^{\frac{1}{2}} A_2^{2,x} \end{pmatrix} \\ &= \begin{pmatrix} \left(Nh^{k}\right)^{\frac{1}{2}} \begin{pmatrix} \sum_i K_i w_{ii} h^2 (x_i - x)' m''(x)(x_i - x)/2 \\ + \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} h^2 (x_i - x)' m''(x)(x_i - x)/2 \end{pmatrix} \\ (Nh^{k+2})^{\frac{1}{2}} \begin{pmatrix} \sum_i K_i w_{ii} h^2 (x_i - x)' m''(x)(x_i - x)/2 \\ + \sum_{i \neq j} \sqrt{K_i}\sqrt{K_j} w_{ij} h^2 (x_i - x)' m''(x)(x_i - x)/2 \end{pmatrix} \end{pmatrix} \\ &\rightarrow \begin{pmatrix} \left(Nh^{k}\right)^{\frac{1}{2}} h^2 \mu_k f(x) W + o_p(1) \\ (Nh^{k+2})^{\frac{1}{2}} O_p(h^2) \end{pmatrix} \\ &= \begin{pmatrix} \left(Nh^{k}\right)^{\frac{1}{2}} h^2 \mu_k f(x) W \\ 0 \end{pmatrix} + o_p(1) \end{aligned}$$

which proves (iii).

Proof of (iv). This proof will examine the variance of each component of $D(N)A^{3,x}$ as well as the covariance between the two components. First,

$$\begin{aligned} VAR\left(\left(Nh^k\right)^{\frac{1}{2}} A_1^{3,x}\right) &= E\left(Nh^k\left(A_1^{3,x}\right)^2\right) \\ &\to f(x)\sigma_\varepsilon^2 \int K^2(\psi)d\psi \frac{1}{N}\sum_i w_{ii} + o(1) \\ &= d_k f(x)\sigma_\varepsilon^2 W + o(1). \end{aligned}$$

Next

$$\begin{aligned} VAR\left(\left(Nh^k\right)^{\frac{1}{2}} A_2^{3,x}\right) &= \sigma_\varepsilon^2 \int f(x+h\psi)K^2(\psi)\psi\psi' d\psi \frac{1}{N}\sum_i w_{ii} \\ &\to v_k f(x)\sigma_\varepsilon^2 W I_k + o(1). \end{aligned}$$

The covariance is shown to go to

$$\begin{aligned} COV\left(\left(Nh^k\right)^{\frac{1}{2}} A_1^{3,x}, \left(Nh^k\right)^{\frac{1}{2}} A_2^{3,x}\right) &= \left(Nh^{k+1}\right) E\left(A_1^{3,x}, A_2^{3,x}\right) \\ &= O(h) \\ &= o(1). \end{aligned}$$

Hence, $VAR\left(D(N)A^{3,x}\right) = V + o(1)$ and since $A^{3,x}$ has a zero mean

$$D(N)A^{3,x} \to N(0,V).$$

Finally, by proving the four statements, we can show that

$$\begin{aligned} D(N)\left(d(x) - \delta(x) - \begin{pmatrix} h^2\mu_k W \\ 0 \end{pmatrix}\right) &= RD(N)\left(A^{2,x} + A^{3,x}\right) - \begin{pmatrix} h^2\mu_k W \\ 0 \end{pmatrix} + o_p(1) \\ &= RD(N)A^{3,x} + o_p(1) \\ &\to R\left(N(0,V)\right) + o_p(1) \\ &\to N(0, RVR) \\ &= N(0, \Sigma_x). \end{aligned}$$

Now we note that $\widehat{\Omega}$ is a consistent estimator of Ω, the proof of it follows from the results of Amemiya (1971) for the parametric model. Thus the asymptotic distribution of $D(N)\left(\widehat{\delta}(x) - \delta(x)\right)$ is the same as that of $D(N)\,(d(x) - \delta(x))$. ∎

Table 1 - Linear Technology ($\gamma = 0$) - Estimates of m

$T = 3$ $\rho = 0.1$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	0.02461	**0.81206**	-0.00294	**0.263450**	0.00101	**0.152292**
$\widehat{m}_{NPFGLS}$	0.03227	1.18472	-0.00918	0.358010	0.02481	0.334819

$T = 5$ $\rho = 0.1$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	-0.00415	**0.62618**	0.00701	**0.262428**	-0.01735	**0.12318**
$\widehat{m}_{NPFGLS}$	0.03685	1.14297	0.00230	0.580311	-0.01315	0.25352

$T = 3$ $\rho = 0.4$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	0.09199	**1.29724**	0.02132	**0.68479**	-0.00176	**0.23748**
$\widehat{m}_{NPFGLS}$	0.04473	1.62509	0.02872	0.95959	0.00232	0.54860

$T = 5$ $\rho = 0.4$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	-0.00965	**1.08460**	0.00787	**0.56886**	0.00764	**0.22496**
$\widehat{m}_{NPFGLS}$	-0.00925	1.62514	0.01551	0.81964	-0.02416	0.34916

$T = 3$ $\rho = 0.8$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	-0.01835	**1.58223**	0.01502	**0.89115**	0.00918	**0.37374**
$\widehat{m}_{NPFGLS}$	0.01257	2.40996	0.04550	1.20262	-0.01792	0.53778

$T = 5$ $\rho = 0.8$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	0.06109	**1.64919**	-0.03037	**0.76996**	0.00668	**0.28034**
$\widehat{m}_{NPFGLS}$	-0.01066	2.40569	-0.01820	1.08072	0.00536	0.44972

Table 2 - Linear Technology ($\gamma = 0$) - Estimates of β

$T = 3$ $\rho = 0.1$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	0.04717	**0.43331**	-0.03045	**0.26559**	0.00498	**0.08858**
$\widehat{\beta}_{NPFGLS}$	0.09091	1.30399	-0.02208	0.74755	0.03207	0.54615

$T = 5$ $\rho = 0.1$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	-0.00855	**0.25194**	0.01213	**0.17773**	-0.00236	**0.04905**
$\widehat{\beta}_{NPFGLS}$	0.06144	2.61958	-0.00354	1.48246	-0.00432	0.74635

$T = 3$ $\rho = 0.4$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	0.03625	**0.46752**	-0.00015	**0.18585**	-0.00420	**0.06197**
$\widehat{\beta}_{NPFGLS}$	-0.03703	1.32680	-0.02840	0.79431	0.00041	0.48636

$T = 5$ $\rho = 0.4$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	0.00386	**0.17615**	0.01756	**0.08143**	0.00871	**0.03467**
$\widehat{\beta}_{NPFGLS}$	0.00991	1.96978	-0.00893	1.08968	-0.04652	0.55036

$T = 3$ $\rho = 0.8$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	0.02128	**0.20419**	-0.01315	**0.08525**	-0.01586	**0.03017**
$\widehat{\beta}_{NPFGLS}$	0.01243	0.77846	-0.00564	0.43128	-0.03533	0.23332

$T = 5$ $\rho = 0.8$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	0.00467	**0.05750**	-0.00071	**0.02854**	0.00137	**0.01069**
$\widehat{\beta}_{NPFGLS}$	-0.04696	0.92974	0.01424	0.52818	-0.00487	0.24382

Table 3 - Quadratic Technology ($\gamma = 2$) - Estimates of m

$T = 3$ $\rho = 0.1$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	3.75215	13.95897	3.78125	14.36996	1.38247	13.83682
$\widehat{m}_{NPFGLS}$	1.25748	**2.259489**	1.01324	**1.487497**	-0.64823	**0.69626**

$T = 5$ $\rho = 0.1$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	3.19402	11.45857	3.21542	11.81433	3.27112	11.80253
$\widehat{m}_{NPFGLS}$	0.60992	**1.52437**	0.45886	**0.80338**	0.329321	**0.42009**

$T = 3$ $\rho = 0.4$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	3.59057	13.24741	3.73497	14.14550	3.69340	13.72657
$\widehat{m}_{NPFGLS}$	1.45033	**2.18517**	0.85695	**1.08555**	0.60328	**0.67794**

$T = 5$ $\rho = 0.4$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	3.21889	12.00314	3.28223	11.51401	3.30367	11.18568
$\widehat{m}_{NPFGLS}$	0.70980	**1.92347**	0.43402	**0.98756**	0.25790	**0.41860**

$T = 3$ $\rho = 0.8$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	3.75286	14.88077	-2.35830	14.53756	3.67594	14.25223
$\widehat{m}_{NPFGLS}$	0.96821	**2.66539**	0.75107	**1.43185**	0.44664	**0.64704**

$T = 5$ $\rho = 0.8$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{m}_{FGLS}$	3.39286	13.66158	3.27061	11.75863	3.25525	11.04261
$\widehat{m}_{NPFGLS}$	0.54470	**2.216837**	0.26445	**1.12261**	0.13944	**0.45376**

Table 4 - Quadratic Technology ($\gamma = 2$) - Estimates of β

$T = 3$ $\rho = 0.1$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	2.76221	8.66174	2.90400	8.97976	2.89316	8.53274
$\widehat{\beta}_{NPFGLS}$	-0.26873	**1.69491**	-0.26373	**0.85174**	-0.36253	**0.53852**

$T = 5$ $\rho = 0.1$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	5.87036	35.40215	5.93808	35.62152	6.05484	37.40021
$\widehat{\beta}_{NPFGLS}$	0.12010	**2.86717**	-0.23052	**1.61403**	-0.51341	**1.21027**

$T = 3$ $\rho = 0.4$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	3.03288	10.38349	3.06586	9.88796	3.12776	10.00722
$\widehat{\beta}_{NPFGLS}$	-0.05755	**1.43259**	-0.30976	**0.88407**	-0.28894	**0.48019**

$T = 5$ $\rho = 0.4$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	5.68197	33.21481	5.92401	35.38354	6.02191	36.17050
$\widehat{\beta}_{NPFGLS}$	0.22978	**2.18748**	-0.24992	**1.28177**	-0.54674	**0.96792**

$T = 3$ $\rho = 0.8$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	3.31846	12.05829	3.44486	12.29401	3.51921	12.52890
$\widehat{\beta}_{NPFGLS}$	0.03322	**0.98089**	-0.08389	**0.51563**	-0.12502	**0.25223**

$T = 5$ $\rho = 0.8$

	$N = 10$		$N = 20$		$N = 50$	
	$Bias$	MSE	$Bias$	MSE	$Bias$	MSE
$\widehat{\beta}_{FGLS}$	5.79115	34.09982	5.91963	35.26351	5.96645	35.51767
$\widehat{\beta}_{NPFGLS}$	0.09283	**0.96495**	-0.23768	**0.72277**	-0.51123	**0.57375**

References

[1] Amemiya, T. (1971). "The Estimation of the Variances in a Variance-Components Model," *International Economic Review*, 12, 1-13.

[2] Baltagi, B. H. (2001). *Econometric Analysis of Panel Data,* West Sussex, John Wiley & Sons.

[3] Baltagi, B. H., Y-J. Chang, and Q. Li (1992). "Monte Carlo Results on Several New and Existing Tests for the Error Component Model," *Journal of Econometrics*, 54, 95-120.

[4] Berg, M. D., Q. Li and A. Ullah. (2000). "Instrumental Variable Estimation of Semiparametric Dynamic Panel Data Models: Monte Carlo Results on Several New and Existing Estimators," *Advances in Econometrics*, 15, 297-315.

[5] Fan, J. and I. Gijbels (1992)."Variable Bandwidth and Local Linear Regression Smoothers," *Annals of Statistics*, 20, 2008-36.

[6] Greene, W. H. (2002). *Econometric Analysis*, Upper Saddle River, Prentice Hall.

[7] Hsiao, C. (2002). *Analysis of Panel Data*, Cambridge, Cambridge University Press.

[8] Kniesner, T. J., and Q. Li. (2002). "Nonlinearity in Dynamic Adjustment: Semiparametric Estimation of Panel Labor Supply," *Empirical Economics*, 27, 131-48.

[9] Li, Q., and J. Wooldridge (2000). "Estimating Semiparametric Econometrics Models by Local Linear Method: With An Application to Gross-Country Growth," *Annals of Economics and Finance*, 1, 337-57.

[10] Maddala, G. S., and T. D. Mount (1973). "A Comparative Study of Alternative Estimators for Variance components Models Used in Econometric Applications," *Journal of the American Statistical Association,* 68, 324-8.

[11] Nerlove, M. (1971). "A Note on Error Components Models," *Econometrica*, 39, 383-96.

[12] Pagan, A., and A. Ullah (1999). *Nonparametric Econometrics*, Cambridge, Cambridge University Press.

[13] Robinson, P. M. (1988). "Root-N-Consistent Semiparametric Regression," *Econometrica,* 56, 931-54.

[14] Taylor, W. E. (1980). "Small Sample Consideration in Estimation from Panel Data," *Journal of Econometrics,* 13, 203-23.