

# A Nonparametric Random Effects Estimator\*

Daniel J. Henderson<sup>†</sup>  
Department of Economics  
State University of New York at Binghamton

and

Aman Ullah<sup>‡</sup>  
Department of Economics  
University of California, Riverside

November 17, 2004

## Abstract

This paper proposes feasible nonparametric random effects estimators. Specifically, we propose feasible versions of the two estimators in Lin and Carroll (2000) and a modified version of the random effects estimator in Ullah and Roy (1998). Further, the consistency properties of these estimators are established.

*Keywords:* Nonparametric, Random Effects, Kernel  
*JEL Classification:* C1, C14, C33

---

\*Acknowledgements: The authors would like to thank Badi Baltagi, Gloria González-Rivera, Jang-Ting Guo, Chihwa Kao, Qi Li, Jeff Racine and an anonymous referee for useful comments. The research on this project has also benefitted from the comments of participants in the University of California, Riverside Econometrics Colloquium, the Syracuse University Econometrics Lunch, the departmental seminar series at Southern Methodist University, and the Eleventh International Conference on Panel Data, College Station, Texas (June 2004).

<sup>†</sup>Corresponding address: Daniel J. Henderson, Department of Economics, State University of New York, Binghamton, NY, 13902, (607) 777-4480, Fax: (607) 777-2681, e-mail: djhender@binghamton.edu.

<sup>‡</sup>Aman Ullah, Department of Economics, University of California, Riverside, CA 92521, (951) 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 data that measure cross-sectional units over a period of time. 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  $i = 1, 2, \dots, N$ ,  $t = 1, 2, \dots, T$ ,  $y_{it}$  is the endogenous variable,  $\alpha$  is the constant term,  $x_{it}$  is a matrix of  $k$  exogenous variables,  $\beta$  is a  $k \times 1$  unknown parameter,  $u_i$  is known as the individual effect and  $v_{it}$  is the random error. 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  $i$ . The above model is also known as the fixed effects or random effects model if  $u_i$  is treated as fixed or random respectively. In this paper, we consider the random effects model, see Baltagi (2001) for details on this and fixed effects models.

Often, the functional form connecting the variables of the model 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. However, there is not much on the nonparametric estimation of panel data models, although see Porter (1996), Ullah and Roy (1998), Lin and Carroll (2000), and Berg et al. (2000).

Lin and Carroll (2000) consider two nonparametric kernel estimators of the unknown regression function and their derivatives. However, their estimators are not a feasible set of estimators since they consider the case of a known variance-covariance matrix of the errors, and they consider only the asymptotic properties of the estimator of the unknown function. In another paper, Ullah and Roy (1998) provide a

feasible estimator, but the asymptotic properties of this estimator are not known. In this paper, we propose feasible versions of the two estimators in Lin and Carroll (2000) and a modified version of the estimator in Ullah and Roy (1998). These three feasible estimators and their consistency properties are presented in Section 2. Section 3 concludes the paper.

## 2 The Model

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

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

where  $i = 1, 2, \dots, N$ ,  $t = 1, 2, \dots, T$ ,  $y_{it}$  is the endogenous variable,  $x_{it}$  is a vector of  $k$  exogenous variables and  $m(\cdot)$  is an unknown smooth function. Further,  $\varepsilon_{it}$  follows the random effects specification

$$\varepsilon_{it} = u_i + v_{it}, \quad (2)$$

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  $v_{jt}$  are uncorrelated for all  $i$ ,  $j$ , and  $t$ .

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

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

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

$$\Omega = \mathbf{I}_N \otimes V. \quad (4)$$

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  $\nabla$  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 finite  $T$ .

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 of errors

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

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  $X_{it} = (1, (x_{it} - x))$ ,  $\delta(x) = (m(x), \beta'(x))'$  is a  $(k+1) \times 1$  vector,  $K(x)$  is an  $NT \times NT$  diagonal matrix of kernel functions  $K(\frac{x_{it}-x}{h})$  and  $h$  is the bandwidth (smoothing) parameter. The estimator obtained is

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

and is called the LLS estimator (see Fan and Gijbels 1992 or Pagan and Ullah 1999 for more details).

The LLS estimator in (6), however, ignores the information contained in the disturbance vector covariance matrix  $\Omega$ . In view of this, we introduce a new estimator, Local Linear Weighted Least Squares (LLWLS), by minimizing

$$(y - X\delta(x))'W(x)(y - X\delta(x)) \quad (7)$$

with respect to  $\delta(x)$ , where  $W(x)$  is a kernel based weight matrix. This provides the kernel estimating equations for  $\delta(x)$  as  $X'W(x)(y - X\delta(x)) = 0$ , which gives

$$d(x) = (X'W(x)X)^{-1}X'W(x)y. \quad (8)$$

We consider the following cases of (8),

$$d_r(x) = (X'W_r(x)X)^{-1}X'W_r(x)y, \quad (9)$$

where  $d_r(x) = (m_r(x), \beta'_r(x))'$ , and for  $r = 1, 2, 3$ ,  $W_1(x) = \sqrt{K(x)}\Omega^{-1}\sqrt{K(x)}$ ,  $W_2(x) = \Omega^{-1}K(x)$  and  $W_3(x) = \Omega^{-\frac{1}{2}}K(x)\Omega^{-\frac{1}{2}}$ . The estimators  $d_1(x)$  and  $d_2(x)$  are as given in Lin and Carroll (2000), and  $d_3(x)$  is as given in Ullah and Roy (1998). We note that when the matrix  $V$ , and hence  $\Omega$ , is a diagonal matrix, then  $W_1(x) = W_2(x) = W_3(x)$ , and hence  $d_1(x) = d_2(x) = d_3(x)$ . Further, in a special case when  $\Omega = \mathbf{I}_{NT}$ ,  $d_1(x) = d_2(x) = d_3(x) = \widehat{\delta}(x)$  in (6). In general, however,  $d_1(x)$ ,  $d_2(x)$ , and  $d_3(x)$  are often different. Further, we notice that the weights of the error  $y - X\delta(x)$  for the estimator  $d_1(x)$  is symmetric, and it essentially applies a LS estimation on the transformed observations  $\Omega^{-\frac{1}{2}}\sqrt{K(x)}y$  and  $\Omega^{-\frac{1}{2}}\sqrt{K(x)}X$ . In view

of this, we refer to  $d_1(x)$  as the local linear nonparametric generalized least squares (NPGLS) estimator. The weights in the estimator  $d_3(x)$  are also symmetric but it amounts to doing LS estimation on the transformed observations  $\sqrt{K(x)}\Omega^{-\frac{1}{2}}y$  and  $\sqrt{K(x)}\Omega^{-\frac{1}{2}}X$ . On the other hand, the weights in  $d_2(x)$  are asymmetric.

Lin and Carroll (2000), for large  $N$  and finite  $T$  and under some regularity condition (see Lin and Carroll 2000, p.523), derive the asymptotic bias and variance of the estimators  $m_r(x)$ ,  $r = 1, 2$ , with scalar regressor ( $k = 1$ ). It is straightforward to extend their results for  $k > 1$ , and also for  $m_3(x)$ . From their results, for  $r = 1, 2, 3$ ,

$$Bias(m_r(x)) = O(h^2), \quad Var(m_r(x)) = O\left(\frac{1}{Nh^k}\right), \quad (10)$$

see Theorems 3 and 4 in Lin and Carroll (2000) with their  $p = 1$  (LLLS) case and  $k = 1$ . Thus, the three estimators given by  $m_r(x)$  are consistent estimators of  $m(x)$  under the assumption that  $h \rightarrow 0$  and  $Nh^k \rightarrow \infty$  as  $N \rightarrow \infty$ . Following Kniesner and Li (2002) and Li and Wooldridge (2000) we can also easily verify that, for  $r = 1, 2, 3$ ,

$$Bias(\beta_r(x)) = O(h^2), \quad Var(\beta_r(x)) = O\left(\frac{1}{Nh^{k+2}}\right). \quad (11)$$

Thus, under the assumption that  $h \rightarrow 0$  and  $Nh^{k+2} \rightarrow \infty$  as  $N \rightarrow \infty$ , the three estimators given by  $\beta_r(x)$  are consistent estimators of  $\beta(x)$ . We note that while Kniesner and Li (2002) results are for the case  $\Omega = \mathbf{I}_{NT}$ , the order of magnitudes of the bias and variance remain unaffected when  $\Omega \neq \mathbf{I}_{NT}$ .

The LLWLS estimator in (8), however, depends upon the unknown parameters  $\sigma_u^2$  and  $\sigma_v^2$ . The spectral decomposition of  $\Omega$  leads to consistent estimators of the variance components as

$$\hat{\sigma}_1^2 = T \sum_t \hat{\varepsilon}_i^2 / N, \quad \hat{\sigma}_v^2 = \frac{1}{N(T-1)} \sum_i \sum_t (\hat{\varepsilon}_{it} - \hat{\varepsilon}_i)^2, \quad (12)$$

where  $\sigma_1^2 \equiv T\sigma_u^2 + \sigma_v^2$ ,  $\hat{\varepsilon}_i = \frac{1}{T} \sum_t \hat{\varepsilon}_{it}$  and  $\hat{\varepsilon}_{it} = y_{it} - \hat{m}(x_{it})$  is the LLLS residual based on the first stage estimator of  $\hat{\delta}(x)$  in (6). The estimate of  $\sigma_u^2$  can be obtained as  $\hat{\sigma}_u^2 = (\hat{\sigma}_1^2 - \hat{\sigma}_v^2) / T$ .

Substituting the estimates of  $\sigma_u^2$  and  $\sigma_v^2$  from (12) into (9) gives a new Local Linear Feasible Weighted Least Squares (LLFWLS) or Nonparametric Feasible Weighted Least Squares (NPFWLS) estimator as

$$\hat{\delta}_r(x) = (X' \hat{W}_r(x) X)^{-1} X \hat{W}_r(x) y, \quad (13)$$

where  $\widehat{W}_r(x)$  is the same as  $W_r(x)$ ,  $r = 1, 2, 3$ , with  $\Omega$  replaced by  $\widehat{\Omega}$ . We note that the feasible estimators  $\widehat{\delta}_r(x)$ ,  $r = 1, 2$ , are not considered in Lin and Carroll (2002) and  $\widehat{\delta}_3(x)$  is an alternative to the Ullah and Roy (1998) feasible estimator.<sup>1</sup> The estimator  $\widehat{\delta}_1(x)$  will be referred to as the feasible NPGLS (NPFGLS) estimator.

For the consistency property of  $\widehat{\delta}_r(x)$  we first note that from (1),  $\widehat{\varepsilon}_{it} = y_{it} - \widehat{m}(x_{it}) = y_{it} - m(x_{it}) - (\widehat{m}(x_{it}) - m(x_{it})) = \varepsilon_{it} + O_p\left((Nh^k)^{-\frac{1}{2}}\right) = \varepsilon_{it} + o_p(1)$  and  $\widehat{\varepsilon}_i = \bar{\varepsilon}_i + o_p(1)$ . Then, following the procedures in the Appendix of Li and Ullah (1998), it can be easily verified that  $\widehat{\sigma}_1^2 = \sigma_1^2 + o_p(1)$ ,  $\widehat{\sigma}_v^2 = \sigma_v^2 + o_p(1)$  and  $\widehat{\sigma}_u^2 = \sigma_u^2 + o_p(1)$ . These give  $\widehat{\Omega} = \Omega + o_p(1)$  and  $\widehat{\delta}_r(x) = \delta_r(x) + o_p(1)$ , and consistency of  $\widehat{\delta}_r(x)$  follows from the consistency of  $d_r(x)$  given above.

### 3 Concluding Remarks

This paper examined the problem of improving the estimation of a one-way random effects error component model. Feasible nonparametric random effects estimators are proposed, and consistency properties of these estimators are established.

---

<sup>1</sup>Note that the Ullah and Roy (1998) feasible estimator is based on the fixed effect (FE) estimator of  $\beta(x)$  for  $k = 1$  as  $\widehat{\beta}(x) = \left(\sum_i \sum_t (x_{it} - \bar{x}_i)^2 K\left(\frac{x_{it}-x}{h}\right)\right)^{-1} \sum_i \sum_t (x_{it} - \bar{x}_i) (y_{it} - \bar{y}_i) K\left(\frac{x_{it}-x}{h}\right)$ , where  $\bar{x}_i = \sum_t x_{it}/T$  and  $\bar{y}_i = \sum_t y_{it}/T$ . However, the actual FE estimator, as shown in Mukherjee and Ullah (2003), is the same as  $\widehat{\beta}(x)$  but  $\bar{x}_i = \sum_t x_{it} K\left(\frac{x_{it}-x}{h}\right)/T$  and  $\bar{y}_i = \sum_t y_{it} K\left(\frac{x_{it}-x}{h}\right)/T$ , which are the weighted means. The asymptotic properties, including consistency, of both of these estimators are not yet known. In view of this, we provide an estimator here by considering a consistent estimator of  $\Omega$ , also see an alternative consistent estimator of  $\Omega$  in Li and Ullah (1998).

## References

- [1] Baltagi, B. H. (2001). *Econometric Analysis of Panel Data*, Wiley and Sons, New York.
- [2] 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.
- [3] Fan, J., and I. Gijbels (1992). "Variable Bandwidth and Local Linear Regression Smoothers," *Annals of Statistics*, 20, 2008-36.
- [4] Kniesner, T. J., and Q. Li. (2002). "Nonlinearity in Dynamic Adjustment: Semiparametric Estimation of Panel Labor Supply," *Empirical Economics*, 27, 131-48.
- [5] Li, Q., and A. Ullah (1998). "Estimating Partially Linear Panel Data Models with One-Way Error Components," *Econometric Reviews*, 17, 145-66.
- [6] Li, Q., and J. Wooldridge (2000). "Estimating Semiparametric Econometrics Models by Local Linear Method: With An Application to Cross-Country Growth," *Annals of Economics and Finance*, 1, 337-57.
- [7] Lin, X., and R. J. Carroll (2000). "Nonparametric Function Estimation for Clustered Data when the Predictor is Measured Without/With Error," *Journal of the American Statistical Association*, 95, 520-34.
- [8] Mukherjee, D., and A. Ullah (2003). "Semiparametric Analysis of Generalized Panel Data: An Application," manuscript, University of California, Riverside.
- [9] Pagan, A., and A. Ullah (1999). *Nonparametric Econometrics*, Cambridge University Press, Cambridge.
- [10] Porter, J. R. (1996). *Essays in Econometrics*, Ph.D. dissertation, MIT.
- [11] Ullah, A., and N. Roy (1998). "Nonparametric and Semiparametric Econometrics of Panel Data," *Handbook of Applied Economics Statistics*, 579-604.