Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP05/2009

Title
Nonparametric Panel Data Models
A Penalized Spline Approach

Author
Gholamreza Hajargasht

Date:
May 2009

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Nonparametric Panel Data Models
A Penalized Spline Approach

Gholamreza Hajargasht

Centre for Efficiency and Productivity Analysis
School of Economics, University of Queensland[1]
and
Department of Agricultural Economics, Shiraz University
May 2009

Abstract

In this paper, we study estimation of fixed and random effects nonparametric panel data models using penalized splines and its mixed model variant. We define a "within" and a "dummy variable" estimator and show their equivalence which can be used as an argument for consistency of the dummy variable estimator when the effects are correlated with regressors. We prove nonparametric counterparts to a variety of the relations between parametric fixed and random effects estimators. Another feature of the approach followed in this paper is the potential to estimate models with heteroscedasticity and autocorrelation in the error term without difficulty. We provide a simulation experiment to illustrate the performance of the estimators.

Keywords: Panel Data, Fixed effects, Random Effects, Nonparametric, Penalized Splines
JEL Codes: C30, D24

---

[1] - Part of this work was undertaken while I was working at the University of Queensland. I would like to thank Chris O'Donnell, Prasada Rao and Bill Griffiths for their comments on the works leading to this paper.

1. Introduction

A number of studies have recently applied non- and semiparametric regression techniques to panel data models (for a review see e.g. Li and Racine 2007). These studies have mainly used kernel smoothing as their underlying nonparametric techniques. In this paper we use penalized splines and its mixed model representations to estimate regression models with panel data.

There are two variants of panel data models: the fixed and the random effects models. In this paper we show how time-invariant fixed and random effects panel data models can be estimated when there are nonparametric elements in the regression function using a penalized spline approach. One important feature of the approach is that we are able to define a "within" and a "dummy variable" estimator and show their equivalence which can be used as an argument for consistency of the dummy variable estimator when the effects are correlated with the regressors. The other feature is that one can use available mixed model softwares to easily estimate the model; furthermore it allows heteroscedasticity and autocorrelation in the effects and error terms which can be a difficulty with most other nonparametric approaches.

The structure of the paper is as follows: Because penalized spline is not a common nonparametric method in econometric literature we provide an overview of univariate nonparametric regression estimation using penalized splines in Section 2. Section 3 deals with the estimation of nonparametric versions of the fixed effects model. Section 4 is devoted to the estimation of the random effects model using penalized least squares and mixed models. In Section 5, we show the relations between our fixed and random effects estimators. Section 6 extends the analysis to the estimation of multivariate models and discusses models allowing heteroscedasticity and serial correlation. The paper concludes with a simulation experiment to illustrate the performance of the proposed estimators.

2- Nonparametric Regression, Penalized Splines and Mixed Models

There are a number of approaches to nonparametric estimation, most of them have been used effectively in a variety of situations and to some extent choice of the method is a matter of taste and experience and sometimes nature of a model or data play a role. In this paper we use penalized splines and one of our objectives is to show it is a desirable approach for estimation of panel data models.

Consider the following regression model[2]

$$y_i = f(x_i) + v_i \qquad (2.1)$$

where f is assumed to be a smooth function. x_i is the only regressor, v_i represents statistical noise and $i = 1, 2, ..., n$ indexes the observations. One way of estimation of such a regression function is to divide the domain of x_i into contiguous intervals and model the relationship between y and x with a separate polynomial in each interval. The dividing points are referred to as knots. The problem with this method is that the estimated function will be discontinuous at knots. This can be overcome by imposing restrictions on the parameters of the polynomials. In practice, $f(x_i)$ in equation (2.1) can be approximated with the polynomial[3]

$$p(x_i) = \beta_0 + \beta_1 x_i + ... + \beta_p x_i^p + \sum_{k=1}^{K} w_k (x_i - \kappa_k)_+^p = \mathbf{x}_i \boldsymbol{\beta} \qquad (2.2)$$

where $(u)_+ = uI(u \geq 0)$, $\kappa_1 < .... < \kappa_K$ are fixed knots, and $\mathbf{x}_i = (1, x_i, ..., x_i^p, (x_i - \kappa_1)_+^p, ..., (x_i - \kappa_K)_+^p)$. Note that $(x_i - \kappa_k)_+^p$ is equal to zero when x_i is smaller than κ_k. We can rewrite the model (2.1) as

---

[2] For an extensive review of penalized spline approach to nonparametric estimation see Ruppert et al. (2003).

[3] In this paper we only discuss polynomial penalized splines, there are other kinds of spline basis for example B-splines and radial splines. For further information on these see e.g. Ruppert et al. (2003) and Eilers and Marx (1996).

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \tag{2.3}$$

where $\mathbf{X}$ is a matrix with $\mathbf{x}_i$ in its *i*-th row, and $\boldsymbol{\beta} = (\beta_0, ..., \beta_p, w_1, ..., w_K)'$. Once the knots have been selected, (2.3) is a linear regression model and can be estimated using ordinary least squares. This method is known as the regression splines and its performance is crucially dependent on the number and location of the knots. Several procedures for selecting the number and locations of the knots are available (see e.g. Smith and Kohn, 1996). The problem with the regression splines is that knot selection procedures are complicated and computationally intensive. In the P-spline approach, we allow the number of knots to be large and fixed (e.g., 20-30 equidistant knots has been found to be adequate for most applications), but to avoid over-fitting (wiggliness) we put a penalty on the w_ks in (2.2) such that

$$\sum_{k=1}^{K} w_k^2 \leq C \tag{2.4}$$

Then the regression spline least squares minimization problem can be written as

$$\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \text{ subject to } \boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} \leq C \tag{2.5}$$

where $\mathbf{K}$ is a diagonal matrix whose first $p+1$ diagonal elements are 0 and the remaining diagonal elements are 1. It can be shown that the penalized least squares minimizer will be

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{K})^{-1}\mathbf{X}'\mathbf{y} \tag{2.6}$$

where λ is a smoothing parameter (the higher the value of λ, the smoother the estimated function will be). The optimal value of λ is usually obtained using a secondary optimization procedure e.g. a cross validation procedure.

Most of what is known about properties of penalized spline estimator has been based on simulation experiments and it has gained popularity in statistical literature due to its numerous successful and easy applications. However, several papers have recently studied asymptotic properties of penalized spline estimators formally (see e.g. Claeskens et al. 2008 and Li and Ruppert 2008). Their basic finding is that with adequate number of

knots and right choice of smoothing parameter, univariate penalized least square estimator can achieve optimal nonparametric convergence rate.
It has also been shown (see e.g. Wand 2003) that the penalization criterion (2.4) can be incorporated into a mixed model framework. To see this, consider a generalized penalized least square problem

$$\min_{\boldsymbol{\beta}_0,\mathbf{w}} (\mathbf{y}-\mathbf{X}_0\boldsymbol{\beta}_0-\mathbf{Zw})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}_0\boldsymbol{\beta}_0-\mathbf{Zw})+\mathbf{w}'\mathbf{G}^{-1}\mathbf{w} \tag{2.7}$$

where $\mathbf{G}$ and $\mathbf{R}$ are two symmetric positive definite matrices. It has been shown (Robinson, 1990, see also Lee and Griffiths 1979) that $\boldsymbol{\beta}_0$ and $\mathbf{w}$ obtained from this minimization is equal to solution to following mixed model.

$$\mathbf{y}=\mathbf{X}_0\boldsymbol{\beta}_0+\mathbf{Zw}+\mathbf{v} \tag{2.8}$$

where $\boldsymbol{\beta}_0$ is an unknown fixed parameter vector to be estimated; $\mathbf{v}$ represents the noise and $\mathbf{w}$ is a random vector satisfying the properties

$$E\begin{bmatrix}\mathbf{w}\\ \mathbf{v}\end{bmatrix}=\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix} \quad \text{and} \quad Cov\begin{bmatrix}\mathbf{w}\\ \mathbf{v}\end{bmatrix}=\begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix} \tag{2.9}$$

Such a model is referred to as a linear mixed model in the statistical literature. Estimation of it can be accomplished by rewriting it in the form

$$\mathbf{y}=\mathbf{X}_0\boldsymbol{\beta}_0+\mathbf{v}^* \quad \text{where} \quad \mathbf{v}^*=\mathbf{Zw}+\mathbf{v} \tag{2.10}$$

This is just a linear model with a generalized covariance matrix

$$\mathbf{V}=Cov(\mathbf{v}^*)=\mathbf{ZGZ}'+\mathbf{R} \tag{2.11}$$

Therefore, $\boldsymbol{\beta}_0$, $\mathbf{V}$ and $\mathbf{w}$ can be estimated\predicted using feasible generalized least squares or maximum likelihood.

Now write the penalized least square problem (2.5) as

$$\min_{\boldsymbol{\beta}} \frac{1}{\sigma_v^2}(\mathbf{y - X\beta})'(\mathbf{y - X\beta}) + \frac{1}{\tau^2}\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} \tag{2.12}$$

where $\lambda = \sigma_v^2 / \tau^2$, then using the above arguments the solution to this problem is also the solution to the following mixed model

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{Zw} + \mathbf{v}$$

$$Cov\begin{bmatrix}\mathbf{v}\\ \mathbf{w}\end{bmatrix} = \begin{pmatrix}\sigma_v^2\mathbf{I}_n & \mathbf{0}\\ \mathbf{0} & \tau^2\mathbf{I}_K\end{pmatrix} \tag{2.13}$$

The mixed model representation of nonparametric regression models has been found to be very useful. It allows nonparametric estimation to be performed using mixed model methodology and software.

3. Nonparametric Fixed Effects Model

In this section, we consider a panel data model in which (i) unobserved heterogeneity is captured by a random term possibly correlated with regressors; (ii) the individual effects are time invariant; (iii) the regression function is nonparametric i.e. of an unknown smooth functional form. Specifically, we assume the relationship between a dependent variable and a single regressor can be represented as follows

$$y_{it} = \alpha_i + f(x_{it}) + v_{it}$$
$$i = 1, \ldots, N \quad t = 1, \ldots, T \tag{3.1}$$

where α_i is an individual specific term (for identification, let $\sum_{i=1}^{N}\alpha_i = 0$), and v_{it} is an error term with mean zero and variance σ_v^2. For simplicity, we assume the panel is balanced, and that there is only one regressor[4].

[4] For extension to multivariate cases see section 6.

Estimation of equation (3.1) using a kernel method has been discussed in Henderson, Carroll and Li (2008). Here we show how (3.1) can be estimated using penalized splines by appealing to either penalized least squares or its mixed model representations.

Similar to the parametric case, we first introduce a "within estimator" which is consistent even when the effects are correlated with regressors and later we show that this estimator is equivalent to a "dummy variable estimator". Define the "within estimator" as follows: Take mean over t in (3.1) to obtain

$$\overline{y}_i = \alpha_i + \frac{1}{T}\sum_{i=1}^{t} f(x_{it}) + \overline{v}_i \tag{3.2}$$

Subtracting (3.2) from (3.1) gives

$$y_i - \overline{y}_i = f(x_{it}) - \frac{1}{T}\sum_{i=1}^{t} f(x_{it}) + v_{it} - \overline{v}_i \tag{3.3}$$

We can write (3.3) in the following regression spline form

$$\mathbf{y} - \overline{\mathbf{y}}_i = (\mathbf{X} - \overline{\mathbf{X}}_i)\boldsymbol{\beta} + \mathbf{v} \tag{3.4}$$

This is a linear regression function which can be estimated by ordinary least squares. However, following the discussion in Section 2, to avoid over-fitting a penalty must be put on the coefficients. The penalized least square "within estimator" of $\boldsymbol{\beta}$ can be obtained as

$$\widehat{\boldsymbol{\beta}}_w = \left\{(\mathbf{X} - \overline{\mathbf{X}}_i)'(\mathbf{X} - \overline{\mathbf{X}}_i) + \lambda\mathbf{K}\right\}^{-1}(\mathbf{X} - \overline{\mathbf{X}}_i)(\mathbf{y} - \overline{\mathbf{y}}_i) \tag{3.5}$$

where λ is a smoothing parameter and its optimal value can be obtained using a secondary optimization procedure (e.g. cross-validation). Because equations 3.3 and 3.4 are independent of the effects, the resulting estimator should have good properties[5].

[5] As it was mentioned in the previous section, studies on asymptotic behavior of penalized spline estimators are at early stages. In this paper we illustrate performance of our estimators using a limited simulation experiment and leave their asymptotic properties for another study.

Now let us define a dummy variable estimator. Write the regression spline form of (3.1) in matrix form as follows

$$\mathbf{y} = \mathbf{D\alpha} + \mathbf{X\beta} + \mathbf{v} \tag{3.6}$$

where $\mathbf{X}$ is as defined above except we have removed the vector of ones from the first column to avoid dummy variable trap. The matrix **D** and vector $\boldsymbol{\alpha}$ are defined as

$$\mathbf{D} = \begin{bmatrix} \mathbf{i} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{i} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{i} \end{bmatrix} \text{ and } \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} \tag{3.7}$$

where $\mathbf{i}$ is a column vector of ones and **0** is a vector of zeros, both of dimension T. By defining $\mathbf{X}^* = (\mathbf{D}, \mathbf{X})$ and $\boldsymbol{\beta}^* = (\boldsymbol{\alpha} \quad \boldsymbol{\beta})'$ we can rewrite (3.6) in the following form

$$\mathbf{y} = \mathbf{X}^* \boldsymbol{\beta}^* + \mathbf{v} \tag{3.8}$$

This is a linear regression function which can be estimated by ordinary least squares. Again to avoid over-fitting a penalty must be put on the spline coefficients in the form of $\sum_{k=1}^{K} w_k^2 \le C$. The penalized least squares estimator $\boldsymbol{\beta}^*$ can then be easily obtained as

$$\boldsymbol{\beta}^* = (\mathbf{X}^{*\prime}\mathbf{X}^* + \lambda^* \mathbf{K}^*)^{-1} \mathbf{X}^{*\prime}\mathbf{y} \tag{3.9}$$

where $\mathbf{K}^*$ is a diagonal matrix with the first $N + p$ diagonal elements equal to zero and the rest equal to one. We may call this a "dummy variable penalized least square" estimator.

We expect the above estimator to have good properties when the effects are fixed parameters but how about when they are random and correlated with regressors? To answer this question we show that the "dummy variable penalized least square" of $\boldsymbol{\beta}$ is equivalent to the "within estimator": To prove this we use a generalized version of Frisch-Waugh theorem.

Generalization of Frisch-Waugh theorem to penalized least square: Consider the following regression spline model (for proof see the appendix).

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{v}$$

Define $\lambda_1, \mathbf{K}_1$ and $\lambda_2, \mathbf{K}_2$ as smoothing parameters and penalty matrix associated with $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively. Then

$$\widehat{\boldsymbol{\beta}}_2 = \left\{\mathbf{X}_2^{'}\mathbf{M}_1\mathbf{X}_2 + \lambda_2\mathbf{K}_2\right\}^{-1}\mathbf{X}_2^{'}\mathbf{M}_1\mathbf{y}$$

where $\mathbf{M}_1 = \left\{\mathbf{I} - \mathbf{X}_1\left(\mathbf{X}_1^{'}\mathbf{X}_1 + \lambda_1\mathbf{K}_1\right)^{-1}\mathbf{X}_1^{'}\right\}$. $\boldsymbol{\beta}_1$ and $\mathbf{M}_2$ are defined similarly.□

For Model (3.6) we have

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{'}\mathbf{M}_1\mathbf{X} + \lambda_2\mathbf{K})^{-1}\mathbf{X}^{'}\mathbf{M}_1\mathbf{y} \tag{3.10}$$

where $\lambda_1 = 0$ and $\mathbf{M}_1 = [\mathbf{I} - \frac{\mathbf{DD'}}{n}]$. $\mathbf{M}_1$ is a mean scaling operator and it is an idempotent matrix. Therefore we can write

$$\widehat{\boldsymbol{\beta}} = \left\{\mathbf{X} - \overline{\mathbf{X}}_i)'(\mathbf{X} - \overline{\mathbf{X}}_i) + \lambda_2\mathbf{K}\right\}^{-1}(\mathbf{X} - \overline{\mathbf{X}}_i)(\mathbf{y} - \overline{\mathbf{y}}_i) \tag{3.11}$$

But this is exactly the within estimator defined in (3.5) if we set $\lambda_2 = \lambda$. The above argument shows that at least with $\lambda_2 = \lambda$, the dummy variable estimator is consistent even if the effects are random and correlated with the error term.

Is it plausible to assume identical λs for both within and dummy variable estimators? The answer is yes. λ is the Lagrangian associated with the penalty constraint and can be obtained from $\hat{\boldsymbol{\beta}}'\mathbf{K}\hat{\boldsymbol{\beta}} = C$. It is reasonable to assume the same penalty constraint and C for both problems since we are trying to estimate the same parameters and the penalty is independent of $\boldsymbol{\alpha}$ for dummy variable model. This leads to the same $\hat{\lambda}$s for both problems because $\hat{\boldsymbol{\beta}}$s obtained from both models are the same. Another argument can be made by appealing to the concept of degree of freedom, which is defined by trace of the hat matrix (i.e. $\mathbf{X}(\mathbf{X'X} + \lambda\mathbf{K})^{-1}\mathbf{X'}$) and is a major way of smoothing parameter selection.

If a particular value say df is specified for degree of freedom of the within model and $N + df$ for degree of freedom of the dummy variable model (because of N extra α_i s in dummy variable model). It can be shown that this choice leads to the same smoothing parameters for both models. Some other criteria like AIC also produce the same smoothing parameter values. However there are criteria like cross-validation which produce different optimal smoothing parameter values.

A third estimator for the fixed effects model can be obtained by appealing to differencing. Such an approach has been followed in Henderson et al. (2008) under a kernel smoothing framework. Here we briefly discuss how differencing can be employed using penalized splines. Note that we can write

$$y_{it} - y_{i1} = f(x_{it}) - f(x_{i1}) + v_{it} - v_{i1} \tag{3.12}$$

to remove the fixed effects (alternatively we can subtract y_{it-1} from y_{it}). Using spline basis we can write the model in the following regression spline form

$$y_{it} - y_{i1} = (\mathbf{x}_{it} - \mathbf{x}_{i1})\boldsymbol{\beta} + \tilde{v}_{it} \tag{3.13}$$

where $\tilde{v}_{it} = v_{it} - v_{it-1}$. As before the above model can be estimated using the penalized least square or its mixed model variant. We can obtain a more efficient estimator by incorporating the following variance-covariance matrix as we see in the next section

$$Cov(\tilde{\mathbf{v}}) = \mathbf{I}_N \otimes \{\sigma_v^2(\mathbf{I}_{T-1} + \mathbf{i}_{T-1}\mathbf{i}'_{T-1})\} \tag{3.14}$$

In this paper our emphasis is on within and dummy variable estimator so we don't discuss difference estimator in any further detail.

The problem of choosing the value of the smoothing parameter by a secondary stage for any of the estimators can be avoided by appealing to mixed model representation of the model. This time we penalize the roughness by assuming $\mathbf{w}$ to be a random vector with mean vector zero and covariance structure

$$Cov(\mathbf{w},\mathbf{v}) = \begin{bmatrix} \sigma_w^2 \mathbf{I}_K & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{NT} \end{bmatrix} \tag{3.15}$$

The model given by (3.4 or 3.6) and (3.15) is in the general format of a mixed model and can be estimated using standard mixed model methodology and software.

4. Nonparametric Random Effects Model

Consider the following model

$$y_{it} = u_i + f(x_{it}) + v_{it} \tag{4.1}$$

where u_i is a time-invariant random variable assumed to be i.i.d. with mean and variance $(0,\sigma_u^2)$ but it is uncorrelated with x_{it}. $v_{it}s$ are the usual random disturbances and are i.i.d. $(0,\sigma_v^2)$. We also accept the standard assumption that v_{it} and u_i are uncorrelated.

Estimation of this model using a kernel smoothing approach has been studied by e.g. Lin and Carroll (2000), Henderson and Ullah (2005) and Su and Ullah (2007). Here we show how the model can be estimated using penalized splines [see also Welsh et al. 2002]. Following the discussion in Section 2 we rewrite (4.1) in its regression spline form

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{Z}\mathbf{w} + \mathbf{u} \otimes \mathbf{i} + \mathbf{v} \tag{4.2}$$

Let $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT}]'$ then rewrite (4.2) as following generalized linear model

$$\mathbf{y} = \mathbf{X}^*\boldsymbol{\beta}^* + \boldsymbol{\varepsilon} \tag{4.3}$$

where $\boldsymbol{\varepsilon} = \mathbf{u} \otimes \mathbf{i} + \mathbf{v}$, $\mathbf{X}^* = (\mathbf{X}_0, \mathbf{Z}), \boldsymbol{\beta}^* = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \mathbf{w} \end{pmatrix}$.

and

$$Cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} = \mathbf{I}_N \otimes \boldsymbol{\Omega} \quad \text{where } \boldsymbol{\Omega} = \sigma_v^2 \mathbf{I}_T + \sigma_u^2 \mathbf{i}_T \mathbf{i}_T' \tag{4.4}$$

where $\mathbf{I}_N$ denotes an identity matrix of dimension N and $\mathbf{i}$ is a column vector of ones. One might think of estimating model (4.4) using a penalized least squares method as explained in Section 3. However, the covariance matrix of $\boldsymbol{\varepsilon}$ is not of an identity form. To obtain a more efficient estimator we must incorporate the information on the structure of the covariance matrix into the estimation process. So we may define a penalized generalized least squares estimator as:

$$\text{Argmin}_{\boldsymbol{\beta}^*,\sigma_u^2,\sigma_v^2} \left\{(\mathbf{y}-\mathbf{X}^*\boldsymbol{\beta}^*)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}^*\boldsymbol{\beta}^*)+\lambda\boldsymbol{\beta}^{*'}\mathbf{K}\boldsymbol{\beta}^*\right\} \qquad (4.5)$$

where λ is a smoothing parameter which controls the smoothness of regression function and can be optimally chosen using e.g. a cross validation criterion. A similar penalized generalized least squares estimator has been proposed by e.g. Wang (1998) in a different context.

Alternatively, we can give a mixed model interpretation to (4.1) by writing

$$\mathbf{w} \sim N(0,\sigma_w^2\mathbf{I}_K) \qquad (4.6)$$

Then we can rewrite (4.1) together with (4.6) in the following form

$$\mathbf{y}=\mathbf{X}_0\boldsymbol{\beta}_0+[\mathbf{Z},\mathbf{D}]\begin{bmatrix}\mathbf{w}\\ \mathbf{u}\end{bmatrix}+\mathbf{v} \qquad (4.7)$$

$$Cov\begin{pmatrix}\mathbf{u}\\ \mathbf{w}\\ \mathbf{v}\end{pmatrix}=\begin{bmatrix}\sigma_u^2\mathbf{I}_N & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \sigma_w^2\mathbf{I}_K & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \sigma_v^2\mathbf{I}_{NT}\end{bmatrix}$$

This is again in the format of a linear mixed model and, consequently, all the components of the model can be estimated using standard mixed model methodology and software.

5- Comparison of Fixed and Random Effects Estimators

Fixed and random effects estimators have been compared in a variety of ways within the parametric context. In this section we show that there are nonparametric counterparts to these results.

First, consider the result that random effects estimator is a within estimator if the observations are "quasi time demeaned" (see e.g. Baltagi 2005). To prove the nonparametric counterpart consider the regression spline form of the nonparametric random effects model (4.3)

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \tag{5.1}$$

where $\boldsymbol{\varepsilon}_i = u_i + \mathbf{v}_i$ and $Cov(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Omega} = \left(\sigma_v^2 \mathbf{I}_T + \sigma_u^2 \mathbf{i}_T \mathbf{i}_T^{'}\right)$

It can be shown that $\boldsymbol{\Omega}^{-1/2} = \frac{1}{\sigma_v}\left[\mathbf{I}_\mathbf{T} - \gamma\, \mathbf{ii}'/T\right]$ where $\gamma = 1 - \left(\frac{\sigma_v^2}{\sigma_v^2 + T\sigma_u^2}\right)^{1/2}$.

Pre-multiplying both sides of (5.1) by $\boldsymbol{\Omega}^{-1/2}$ we obtain

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{X}}_i \boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_i \text{ where } Cov(\tilde{\boldsymbol{\varepsilon}}_i) = \sigma_u^2 \mathbf{I}_T \tag{5.2}$$

The *t*th element of $\tilde{\mathbf{y}}_i$ is $y_{it} - \gamma \bar{y}_i$ and similarly for $\tilde{\mathbf{X}}_i$. Therefore we can write

$$y_{it} - \gamma \bar{y}_i = (\mathbf{x}_{it} - \gamma \bar{\mathbf{x}}_i)\boldsymbol{\beta} + \varepsilon_{it} - \gamma \bar{\varepsilon}_i \tag{5.3}$$

This shows that similar to parametric case when $T\sigma_u^2/\sigma_v^2$ is large, γ becomes close to one. In fact, $\gamma \to 1$ as $T \to \infty$ or $\sigma_u^2/\sigma_v^2 \to \infty$. For large *T*, estimates from fixed effects and random effects are similar but even with small *T*, random effects is close to fixed effects if the estimated variance of u_i is large relative to the estimated variance of v_{it} as it is the case in many applications.

Now consider the result that parametric random effects estimator is a weighted average of between and within estimator (see e.g. Hsiao 2003, pp 35-37 or Baltagi 2005, pp 18). To show the nonparametric counterpart, write the model as

$$\mathbf{y} = \mu + \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{5.4}$$

where here **X** does not include the vector of ones. Using the Corollary (1) in the appendix we can write

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left\{\mathbf{X}'\mathbf{M}_1\mathbf{X} + \lambda\mathbf{K}\right\}^{-1}\mathbf{X}'\,\mathbf{M}_1\mathbf{y} \tag{5.5}$$

where $\mathbf{M}_1 = \boldsymbol{\Sigma}^{-1/2}\left\{\mathbf{I} - \boldsymbol{\Sigma}^{-1/2}\mathbf{i}_{NT}\left(\mathbf{i}'_{NT}\boldsymbol{\Sigma}^{-1}\mathbf{i}_{NT}\right)^{-1}\mathbf{i}'_{NT}\boldsymbol{\Sigma}^{-1/2}\right\}\boldsymbol{\Sigma}^{-1/2}$

It can be shown with $\boldsymbol{\Sigma}$ defined in (4.4), $\mathbf{M}_1$ becomes

$$\mathbf{M}_1 = \frac{1}{\sigma_u^2}\left\{\mathbf{I_N} \otimes \mathbf{Q} + \psi\left(\mathbf{I_N} - \mathbf{i_N}\mathbf{i}'_\mathbf{N}/N\right) \otimes \mathbf{P}\right\} \tag{5.6}$$

$$\mathbf{Q} = \mathbf{I} - \frac{1}{T}\mathbf{i}_T\mathbf{i}'_T\ ,\ \mathbf{P} = \frac{1}{T}\mathbf{i}_T\mathbf{i}'_T \text{ and } \psi = \frac{\sigma_v^2}{\sigma_v^2 + T\sigma_u^2}$$

With $\mathbf{M}_1$ defined as above we can write

$$\widehat{\boldsymbol{\beta}}_{GLS} = \left\{\mathbf{X}'\left[\mathbf{I}_N \otimes \mathbf{Q} + \psi(\mathbf{I}_N - \mathbf{i_N}\mathbf{i}'_\mathbf{N}/N) \otimes \mathbf{P}\right]\mathbf{X} + \lambda\mathbf{K}\right\}^{-1} \tag{5.7}$$

$$\times\left\{\mathbf{X}'(\mathbf{I}_N \otimes \mathbf{Q})\mathbf{y} + \psi\mathbf{X}'\left[(\mathbf{I}_N - \mathbf{i_N}\mathbf{i}'_\mathbf{N}/N) \otimes \mathbf{P}\right]\mathbf{y}\right\}$$

This can be rewritten as a weighted average of a between and a within estimator

$$\widehat{\boldsymbol{\beta}}_{GLS} = \boldsymbol{\Delta}\widehat{\boldsymbol{\beta}}_b + (\mathbf{I}_K - \boldsymbol{\Delta})\widehat{\boldsymbol{\beta}}_w \tag{5.8}$$

if we write $\lambda = \lambda_1 + \psi\lambda_2$ [6] and

$$\boldsymbol{\Delta} = \psi\left\{\mathbf{X}'\left[\mathbf{I}_N \otimes \mathbf{Q} + \psi(\mathbf{I}_N - \mathbf{i_N}\mathbf{i}'_\mathbf{N}/N) \otimes \mathbf{P}\right]\mathbf{X} + \lambda\mathbf{K}\right\}^{-1}$$

$$\times\left(\mathbf{X}'\left[(\mathbf{I}_N - \mathbf{i_N}\mathbf{i}'_\mathbf{N}/N) \otimes \mathbf{P}\right]\mathbf{X} + \lambda_2\mathbf{K}\right) \tag{5.9}$$

$$\widehat{\boldsymbol{\beta}}_b = \left\{\mathbf{X}'\left[(\mathbf{I}_N - \mathbf{i_N}\mathbf{i}'_\mathbf{N}/N) \otimes \mathbf{P}\right]\mathbf{X} + \lambda_2\mathbf{K}\right\}^{-1}\mathbf{X}'\left[(\mathbf{I}_N - \mathbf{i_N}\mathbf{i}'_\mathbf{N}/N) \otimes \mathbf{P}\right]\mathbf{y}$$

$$\widehat{\boldsymbol{\beta}}_w = \left\{\mathbf{X}'\left[\mathbf{I}_N \otimes \mathbf{Q}\right]\mathbf{X} + \lambda_1\mathbf{K}\right\}^{-1}\mathbf{X}'\left[\mathbf{I}_N \otimes \mathbf{Q}\right]\mathbf{y}$$

---

[6] Our experiments have shown that always $\lambda > \lambda_1$ and therefore λ_2 is well-defined although we haven't been able to formally prove it.

Finally, consider an extension of Mundlak (1978) model where $y_{it} = u_i + f(x_{it}) + v_{it}$ and u_i is correlated with x_{it} through following relation

$$u_i = \frac{1}{T}\sum_{i=1}^{T} g(x_{it}) + \omega_i \tag{5.10}$$

Note that we are also extending the correlation structure in Mundlak model from a linear form to a nonparametric form. We can rewrite the model as

$$y_{it} = \frac{1}{T}\sum_{t=1}^{T} g(x_{it}) + f(x_{it}) + \omega_i + v_{it} \tag{5.11}$$

If we choose the same knots for g and f we can write

$$\mathbf{y} = \mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} + \mathbf{P}_1\mathbf{X}(\boldsymbol{\beta} + \mathbf{a}) + \boldsymbol{\varepsilon} \tag{5.12}$$

where $\mathbf{P}_1 = \mathbf{I}_N \otimes \mathbf{i}_\mathbf{T}\mathbf{i}_\mathbf{T}' / T$ and $\mathbf{Q}_1 = \mathbf{I}_{NT} - \mathbf{P}_1$. It is easy to see that $\mathbf{X}'\mathbf{Q}_1\boldsymbol{\Sigma}^{-1}\mathbf{P}_1\mathbf{X} = \mathbf{0}$ therefore we can use the corollary 2 in the appendix to write

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^*_{GLS} &= \left\{\mathbf{X}'\,\mathbf{Q}_1\boldsymbol{\Sigma}^{-1}\mathbf{Q}_1\mathbf{X} + \lambda_1\mathbf{K}\right\}^{-1}\mathbf{X}'\,\mathbf{Q}_1\boldsymbol{\Sigma}^{-1}\mathbf{Q}_1\mathbf{y} \\ &= \left\{\mathbf{X}'\,\mathbf{Q}_1\mathbf{X} + \lambda_1\mathbf{K}\right\}^{-1}\mathbf{X}'\,\mathbf{Q}_1\mathbf{y} = \widehat{\boldsymbol{\beta}}_W\end{aligned} \tag{5.13}$$

and

$$\left(\widehat{\boldsymbol{\beta}}^* + \widehat{\mathbf{a}}\right)_{GLS} = \left\{\mathbf{X}'\,\mathbf{P}_1\boldsymbol{\Sigma}^{-1}\mathbf{P}_1\mathbf{X} + \lambda_2\mathbf{K}_1\right\}^{-1}\mathbf{X}'\,\mathbf{P}_1\boldsymbol{\Sigma}^{-1}\mathbf{P}_1\mathbf{y}$$

$$= \left\{\mathbf{X}'\,\mathbf{P}_1\mathbf{X} + \lambda_2\mathbf{K}_1\right\}^{-1}\mathbf{X}'\,\mathbf{P}_1\mathbf{y} = \widehat{\boldsymbol{\beta}}_b$$

$$\Rightarrow \widehat{\mathbf{a}}_{GLS} = \widehat{\boldsymbol{\beta}}_b - \widehat{\boldsymbol{\beta}}^*_{GLS} = \widehat{\boldsymbol{\beta}}_b - \widehat{\boldsymbol{\beta}}_w \tag{5.14}$$

Combining (5.14) and (5.8) we obtain

$$\widehat{\boldsymbol{\beta}}_{GLS} = \boldsymbol{\Delta}\widehat{\mathbf{a}} + \widehat{\boldsymbol{\beta}}_W$$

Pre-multiplying both sides by $\mathbf{X}$ we obtain

$$\mathbf{X}\widehat{\boldsymbol{\beta}}_{GLS} = \mathbf{X}\boldsymbol{\Delta}\widehat{\mathbf{a}} + \mathbf{X}\widehat{\boldsymbol{\beta}}_W \Rightarrow \widehat{\mathbf{f}}_{GLS} = \widehat{\mathbf{f}}_W + \mathbf{X}\boldsymbol{\Delta}\widehat{\mathbf{a}} \tag{5.15}$$

We expect $\widehat{\mathbf{f}}_W$ to be asymptotically unbiased but with T fixed and $N \to \infty$ term $\mathbf{X}\boldsymbol{\Delta}\widehat{\mathbf{a}}$ doesn't converge to zero and therefore $\widehat{\mathbf{f}}_{GLS}$ is asymptotically biased. With $T \to \infty$ $\mathbf{X}\boldsymbol{\Delta}\widehat{\mathbf{a}}$ converges to zero (because ψ goes to zero) and $\widehat{\mathbf{f}}_{GLS}$ tends to $\widehat{\mathbf{f}}_W$.

6. Extensions

The above analysis can be extended in many ways: First, consider extension to multivariate cases. The Multivariate model can be written as

$$y_{ii} = f(x_{1it}, x_{2it}, ..., x_{dit}) + u_i + v_{it} \qquad (6.1)$$

where $(x_1, x_2, ..., x_d)'$ is a vector of regressors and f is a smooth function containing some nonparametric components. It can be of partially linear, additive or a fully nonparametric form. The first step for estimation of model (6.1) under a penalized spline framework is to derive the regression spline equivalent of the model and to define the penalty matrix. Write the regression spline form of the model as

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{u} \otimes \mathbf{I}_T + \mathbf{v} \qquad (6.2)$$

where $\mathbf{X} = [\mathbf{X}_0, \mathbf{Z}_1, ...\mathbf{Z}_l]$ and $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_0' & \mathbf{w}_1' & \dots & \mathbf{w}_l' \end{bmatrix}'$. l can be different from d (see below). Corresponding to $\mathbf{X}$ and $\boldsymbol{\beta}$ define a penalty matrix with following block-diagonal structure

$$\mathbf{K} = \begin{bmatrix} \mathbf{0}_{p_0} & \cdots & & \mathbf{0} \\ \vdots & \lambda_1 \mathbf{I}_{K_1} & & \vdots \\ & & \ddots & \\ \mathbf{0} & \cdots & & \lambda_l \mathbf{I}_{K_l} \end{bmatrix} \qquad (6.3)$$

Assuming linear splines if f is univariate we have $\mathbf{X} = [\mathbf{X}_0, \mathbf{Z}_1]$ *and* $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_0 & \mathbf{w}_1 \end{bmatrix}'$, $\mathbf{K}$ is consist of two block $\mathbf{0}_{p_0}$ and $\lambda_1 \mathbf{I}_{K_1}$ where p_0 (the dimension of $\boldsymbol{\beta}_0$) is equal to 2, we only have one λ parameter, and K_1 (the dimension of $\mathbf{w}_1$) is equal to the number of knots. If f is a fully nonparametric multivariate function, every thing is the same except that p_0 is equal to the number of regressors plus one and in **Z** matrix we have multivariate splines. If extra variables are added to the model in a linear parametric fashion (making the model partially linear), **K**, $\boldsymbol{\beta}$ and **X** still have the same structure but p_0 is equal to the number of regressors in nonparametric form plus the number of variables in the linear form plus one. If some variables are added in a nonparametric additive manner, we can have as many λs as there are additive functions, and K_i $(i = 1, 2, ..., l)$ is the number of knots associated with the *i*-th variable in the additive form. We don't go in any further detail, to learn more about multivariate estimation see e.g. Ruppert et al. 2003. What is worth noting here is that all the results obtained in previous sections apply to multivariate cases as well.

Another feature of the penalized spline approach is that we can allow for a variety of heteroscedasticity and autocorrelation structures in the error term within the same framework and interestingly such an approach has been shown to have desirable properties. A series of studies have shown that the presence of correlation in the error term can have serious effects on nonparametric estimation. For example, Lin and Carroll (2000) found that first generation kernel-based estimators incorporating covariance matrix information are generally asymptotically less efficient than estimates from a model ignoring the correlation structure. Welsh et al. (2002) studied this in more detail and proposed a new kernel approach which is more efficient than standard kernel estimation ignoring the correlation structure, but the Welsh et al estimator is still significantly less efficient than the GLS spline estimator. More recently, Xiao et al. (2003) showed that a modified kernel-based approach proposed by Wang (2003) is as efficient as the GLS spline estimator.

It is well-known that in the presence of correlated errors, standard smoothing parameter selectors fail to work (Altman, 1990, Hart, 1991 or Opsomer et al. 2001). The problem can be avoided by taking the correlation structure explicitly into account as it has been done by Wang (1998) for splines and in Altman (1990) and Hart (1991) for kernels. However, the correlation structure is typically unknown and even small misspecification of the correlation structure can result in serious over- or under–fitting, as demonstrated in Opsomer et al. (2001). Recently, Krivobokova and Kauermann (2007) provided both theoretical and simulation evidence that that a maximum-likelihood-based choice of the smoothing parameter (i.e. mixed model) is very robust against a misspecified correlation structure, and over-fitting is circumvented even for errors that are strongly correlated.

In summary, both standard panel data models and those with more general covariance structures can be easily estimated using penalized splines and interestingly, studies in other contexts have shown that the resulting estimators are at least as good as other approaches.

7. A Simulation Experiment

In this section, we use a Monte Carlo simulation to study performance of the proposed estimators. We follow Wang (2003) and Henderson et al. (2008)[7] to generate the following data generating process: $y_{it} = Sin(2\pi x_{it}) + u_i + v_{it}$ where x_{it} is i.i.d. $Unif[-1,1]$, and v_{it} is i.i.d. Nor(0,1). Let v_{it} denote an i.i.d. $Unif[-1,1]$ sequence of random variables. We generate $u_i = v_i + c\bar{x}_i$, where $\bar{x}_i = \sum_{t=1}^{T} x_{it} / T$. $c = 0$ gives the random effects and $c = 1$ gives the fixed effects model. Note that x_{it} and u_i are correlated for the fixed effects model and uncorrelated for the random effects. The variances of v_{it} and v_i are set to one.

We employed random effects, dummy variable and within estimators to estimate regression function *f*. We used a mixed model approach to estimate all the models, therefore there was no need to use a secondary procedure to choose smoothing parameters. To assess the performance of the estimators we used average mean squared error (AMSE) criterion defined by

$$AMSE = \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\hat{f}(x_{it,m}) - f(x_{it,m}) \right]^2 \Big/ NTM$$

where the subscript *m* denotes the *m*th replication. In each experiment we use M = 1000 replications. The number of time periods (t) is set at 3, while the number of cross-sections

[7] Henderson et al. (2008) consider $y_{it} = Sin(2x_{it}) + u_i + v_{it}$

(N) is varied between 50, 100 and 200. The estimation results can be seen in Table 1 and they can be summarized as follows

• When c = 0 i.e. the data generating process is that of a random effects model, we see that the random effects estimator is more efficient (it has a smaller AMSE than the fixed effects estimators). We also see that dummy variable and within estimators gives the same results when the same smoothing parameters are used for both models and they gives similar numbers when smoothing parameters are set according to a maximum likelihood criteria. Also as expected, for all estimators, the AMSE decreases as N becomes larger.

• When *c =1* i.e. the data generating process is that of a fixed effects model and x_{it} and u_i are correlated, we expect that the fixed effects estimator to be consistent but random effects estimator to be inconsistent. As we see from the table, the within and the dummy variable estimator provide results comparable to the previous case and AMSE gets smaller with similar rates when N increases. However, AMSE associated with the random effects estimator is larger and the bias doesn't seem to converge to zero. To illustrate this, we calculated AMSE of random effects estimator for different values of N from 50 to 5000. The results have been depicted in Graph 1. As we see, AMSE of random effects model (*c=0*) seems to converge to zero but AMSE of fixed effects model (*c=1*) seems to converge to some nonzero value.

• We also depicted AMSE of the random effect estimator adjusting AMSE of the estimator of the model with N=50 by optimal nonparametric convergence rate [$(NT)^{-4/5}$)]. The resulting graph closely follows the graph obtained from the simulation experiment. This limited experiment suggests that the random effects estimator might actually achieve the optimal convergence rate (a similar result can be obtained for fixed effects estimator).

Conclusion

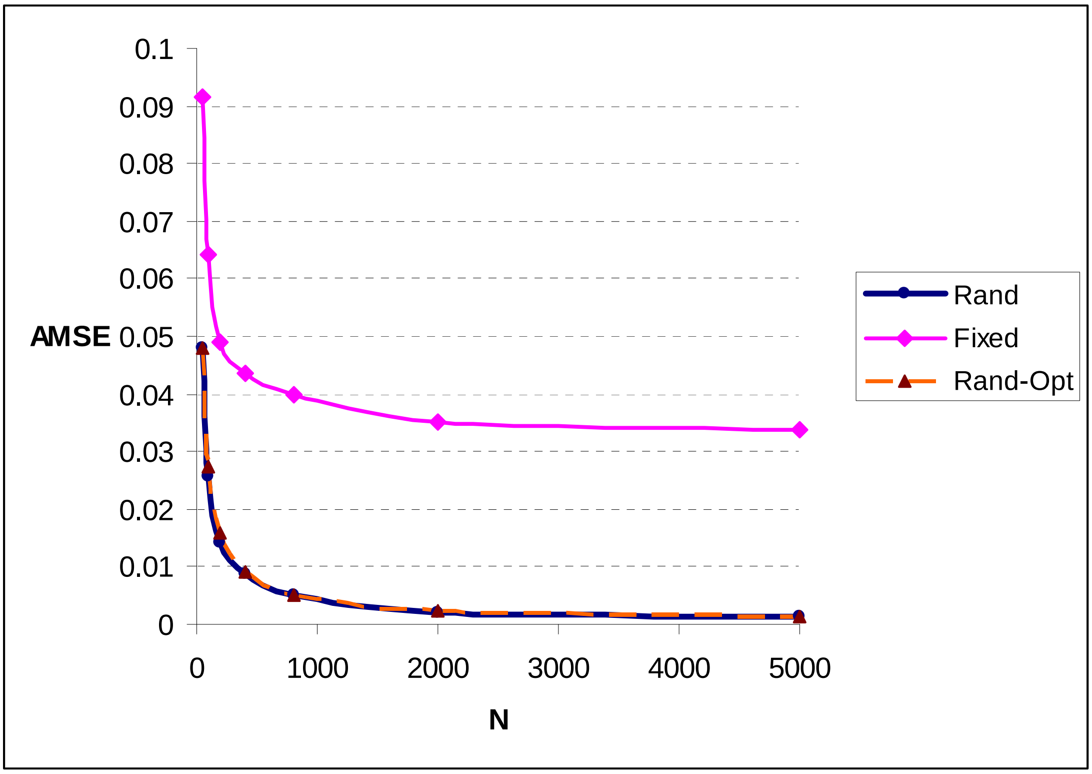
In this paper, we showed how penalized splines can be employed to estimate fixed and random effects panel data models with nonparametric components. It was shown that, under this framework, we can define a within and a dummy variable estimator for the fixed effects model and proved that the dummy variable estimator is equivalent to the "within estimator" and therefore consistent when the regressors are correlated with the effects. It was also shown how random effects models can be estimated and proved that a variety of proven relationships between parametric fixed and random models also holds for our nonparametric estimators. A Monte Carlo experiment was employed to illustrate the performance of the estimators.

Table 1: Average mean squared errors of the random effects, dummy variable and within estimators when the data generation process is a random effects model and when it is a fixed effects model.

	Random Effects Model			**Fixed Effects Model**		
No. Observation	50	100	200	50	100	200
Random Effects	0.04792	0.02562	0.01426	0.09143	0.06405	0.04886
Within	0.06254	0.03546	0.01903	0.06254	0.03546	0.01903
Dummy[1]	0.06115	0.03422	0.01808	0.06115	0.03422	0.01808
Dummy[2]	0.06254	0.03546	0.01903	0.06254	0.03546	0.01903

Note: "Within" refers to within estimator, "Dummy[1]" refers to dummy variable estimator when the smoothing parameter is set equal to that of within estimator and "Dummy[2]" refers to dummy variable estimator when the smoothing parameter has been selected automatically. All the estimations have been done using a mixed model approach. For AMSE calculations we have discarded the 10 lowest and highest x values to avoid boundary effects. The number of Monte Carlo replications is 1000.

Graph 1. AMSE of Random Effects Estimator



N represents the number of cross sections. "Rand" and "Fixed" represent random effects and fixed effects models respectively. "Rand-Opt" depicts AMSE of the random effects model adjusted by optimal nonparametric convergence rate.

References

Altman, N. "Kernel Smoothing of Data with Correlated Errors." *Journal of the American Statistical Association*, 1990, 85, 749-59.

Baltagi, B.H. "*Econometric Analysis of Panel Data.*" Third Edition, John Wiley & Sons, 2005

Claeskens, G., T. Krivobokova and J. Opsomer "Asymptotic Properties of Penalized Spline Estimators." *Biometrika*, 2008, 1-26

Eilers, P., and B. Marx. "Flexible Smoothing with B-Splines and Penalties (with Discussion)." *Statistical Science*, 1996, 11, 89-121.

Hajargasht, G. *Non- and Semiparametric Stochastic Frontiers: A Penalized Spline Approach*, Unpublished PhD Thesis, University of Queensland, 2006.

Hart, J.D. "Kernel Regression Estimation with Time Series Errors." J*ournal of the Royal Statistical Society (Series B)*, 1991, 53,173-87.

Hsiao, C. *Analysis of Panel Data*. 2nd editions, Cambridge, UK, Cambridge University Press, 2003

Henderson, D., and A. Ullah. "A Nonparametric Random Effects Estimator" *Economics Letters*, 2005, 88, 403-407.

Henderson, D., R. Carroll and Q. Li, "Nonparametric estimation and testing of fixed effects panel data models." *Journal of Econometrics*, 2008, 144, 257–275.

Krivobokova, T., Kauermann, G. "A Note on Penalized Spline Smoothing with Correlated Errors" *Journal of the American Statistical Association*, 2007, 102(480): 328-1337

Lee, L. and W. Griffiths. "The Prior Likelihood and Best Linear Unbiased Prediction in Stochastic Coefficient Linear Models." Working Paper No. 1, Department of Econometrics, University of New England, Armidale, 1979,

Li, Q. and J.S. Racine. *Nonparametric Econometrics: Theory and Practice.* Princeton University Press, 2007

Li, Y. and D. Ruppert. "On the asymptotics of penalized splines." *Biometrika,* 2008, 95, 415–436.

Lin, X., and R. Carroll. "Semiparametric Regression for Clustered Data Using Generalized Estimating Equations." *Journal of American Statistical Association*, 2000, 96, 1045-1056.

Lin, X., A. Wang, A. Welsh and R. Carroll. "Equivalent Kernels of Smoothing Splines in Nonparametric Regression for Clustered/Longitudinal Data." *Biometrika*, 2004, *91*, 177-93.

Mundlak, Y. "On the Pooling of Time Series and Cross Section Data." *Econometrica*, 1978, 46, 69–85.

Opsomer, J., Y.Wang, and Y. Yang. "Nonparametric Regression with Correlated Errors." *Statistical Science*, 2001, 34, 134-53.

Robinson, G. "That BLUP Is a Good Thing." *Statistical Science*, 1990, *6*, 15-51.

Ruppert, D., M. Wand and R. Carroll. *Semiparametric Regression*. New York: Cambridge University Press, 2003.

Smith, M., and Kohn, R. "Nonparametric Regression via Bayesian Variable Selection." *Journal of Econometrics*, 1996, 75, 317-44.

Su, L. and A. Ullah. "More Efficient Estimation of Nonparametric Panel Data Models with Random Effects." *Economics Letters*, 2007, 96 (3), 375-380

Ullah, A. and Roy, N. "Nonparametric and Semiparametric Econometrics of Panel Data." in *Handbook of Applied Economics Statistics*, A. Ullah and D. E. A. Giles, eds., Marcel Dekker, New York, 579-604.

Wand, M. "Smoothing and Mixed Models." *Computational Statistics*, 2003, 18, 223-49.

Wang, Y. "Smoothing Spline Models with Correlated Random Errors." *Journal of the American Statistical Association*, 1998, 93, 34-348.

Wang, N. "Marginal Nonparametric Kernel Regression Accounting for within-Subject Correlation." *Biometrika*, 2003, 90, 43-52.

Welsh, A., X., Lin, X. and R. Carroll. "Marginal Longitudinal Nonparametric Regression: Locality and Efficiency of Spline and Kernel Methods." *Journal of the American Statistical Association*, 2002, 97, 482-493.

Xiao, Z., O. Linton, R. Carroll and E. Mammen. "More Efficient Kernel Estimation in Nonparametric Regression with Autocorrelated Errors." *Journal of the American Statistical Association*, 2003, 98, 980-992.

Appendix 1

Generalization of Frisch-Waugh theorem to penalized least square: Consider the following regression spline model

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{v} \tag{A.1}$$

Define $\lambda_1, \mathbf{K}_1$ and $\lambda_2, \mathbf{K}_2$ as smoothing parameters and penalty matrix associated with $\mathbf{X}_1$ and $\mathbf{X}_2$ then

$$\widehat{\boldsymbol{\beta}}_2 = \left\{\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2 + \lambda_2\mathbf{K}_2\right\}^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y} \text{ and}$$

where $\mathbf{M}_1 = \left\{\mathbf{I} - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1 + \lambda_1\mathbf{K}_1\right)^{-1}\mathbf{X}_1'\right\}$. $\boldsymbol{\beta}_1$ and $\mathbf{M}_2$ are defined similarly.

Proof: Penalized least square estimator associated with (A.1) with differing smoothing parameters and penalty matrix for $\mathbf{X}_1$ and $\mathbf{X}_2$ can be written as

$$\left(\begin{bmatrix}\mathbf{X}_1' \\ \mathbf{X}_2'\end{bmatrix}\begin{bmatrix}\mathbf{X}_1', \mathbf{X}_2'\end{bmatrix} + \begin{bmatrix}\lambda_1\mathbf{K}_1 & \mathbf{0} \\ \mathbf{0} & \lambda_2\mathbf{K}_2\end{bmatrix}\right)\begin{bmatrix}\widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y}\end{bmatrix} \tag{A.2}$$

Performing the multiplications we obtain

$$\begin{pmatrix}\left(\mathbf{X}_1'\mathbf{X}_1 + \lambda_1\mathbf{K}_1\right)\widehat{\boldsymbol{\beta}}_1 + \mathbf{X}_1'\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2 \\ \mathbf{X}_2'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1 + \left(\mathbf{X}_2'\mathbf{X}_2 + \lambda_2\mathbf{K}_2\right)\widehat{\boldsymbol{\beta}}_2\end{pmatrix} = \begin{bmatrix}\mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y}\end{bmatrix} \tag{A.3}$$

Pre-multiply both sides of the upper row by $\mathbf{X}_2'\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1 + \lambda_1\mathbf{K}_1\right)^{-1}$ to obtain

$$\begin{pmatrix}\mathbf{X}_2'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1 + \mathbf{X}_2'\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1 + \lambda_1\mathbf{K}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2 \\ \mathbf{X}_2'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1 + \left(\mathbf{X}_2'\mathbf{X}_2 + \lambda_2\mathbf{K}_2\right)\widehat{\boldsymbol{\beta}}_2\end{pmatrix} = \begin{bmatrix}\mathbf{X}_2'\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1 + \lambda_1\mathbf{K}_1\right)^{-1}\mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y}\end{bmatrix} \tag{A.4}$$

Subtracting the lower row from the second and solve for $\widehat{\boldsymbol{\beta}}_2$ we obtain

$$\widehat{\boldsymbol{\beta}}_2 = \left\{\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2 + \lambda_2\mathbf{K}_2\right\}^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y} \tag{A.5}$$

where $\mathbf{M}_1 = \left\{\mathbf{I} - \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1 + \lambda_1\mathbf{K}_1\right)^{-1}\mathbf{X}_1'\right\}$

Corollary (1) If $Cov(\mathbf{v}) = \sigma_v^2 \mathbf{V}$ the above theorem applies but $\mathbf{M}_1$ should be defined as follows

$$\mathbf{M}_1 = \mathbf{V}^{-1/2} \left\{ \mathbf{I} - \mathbf{V}^{-1/2} \mathbf{X}_1 \left(\mathbf{X}_1' \mathbf{V}^{-1} \mathbf{X}_1 + \lambda_1 \mathbf{K}_1 \right)^{-1} \mathbf{X}_1' \mathbf{V}^{-1/2} \right\} \mathbf{V}^{-1/2} \quad \text{(A.6)}$$

This can be easily proven using transformation of $\mathbf{X}_1^* = \mathbf{V}^{-1/2} \mathbf{X}_1$ and $\mathbf{y}^* = \mathbf{V}^{-1/2} \mathbf{y}$ and apply the theorem.

Corollary (2) If $Cov(\mathbf{v}) = \sigma_v^2 \mathbf{V}$ and $\mathbf{X}_1' \mathbf{V}^{-1} \mathbf{X}_2 = \mathbf{0}$ then

$$\hat{\boldsymbol{\beta}}_2 = \left\{ \mathbf{X}_2' \mathbf{V}^{-1} \mathbf{X}_2 + \lambda_2 \mathbf{K}_2 \right\}^{-1} \mathbf{X}_2' \mathbf{V}^{-1} \mathbf{y} \quad \text{and} \quad \hat{\boldsymbol{\beta}}_1 = \left\{ \mathbf{X}_1' \mathbf{V}^{-1} \mathbf{X}_1 + \lambda_1 \mathbf{K}_1 \right\}^{-1} \mathbf{X}_1' \mathbf{V}^{-1} \mathbf{y}$$

To prove this notice that with $Cov(\mathbf{v}) = \sigma_v^2 \mathbf{V}$ (A.3) becomes

$$\begin{pmatrix} \left(\mathbf{X}_1' \mathbf{V}^{-1} \mathbf{X}_1 + \lambda_1 \mathbf{K}_1 \right) \hat{\boldsymbol{\beta}}_1 + \mathbf{X}_1' \mathbf{V}^{-1} \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 \\ \mathbf{X}_2' \mathbf{V}^{-1} \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 + \left(\mathbf{X}_2' \mathbf{V}^{-1} \mathbf{X}_2 + \lambda_2 \mathbf{K}_2 \right) \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{bmatrix} \mathbf{X}_1' \mathbf{V}^{-1} \mathbf{y} \\ \mathbf{X}_2' \mathbf{V}^{-1} \mathbf{y} \end{bmatrix} \quad \text{(A.7)}$$

Since we have assumed $\mathbf{X}_1' \mathbf{V}^{-1} \mathbf{X}_2 = \mathbf{0}$ we have (A.8) which proves the theorem

$$\begin{pmatrix} \left(\mathbf{X}_1' \mathbf{V}^{-1} \mathbf{X}_1 + \lambda_1 \mathbf{K}_1 \right) \hat{\boldsymbol{\beta}}_1 \\ \left(\mathbf{X}_2' \mathbf{V}^{-1} \mathbf{X}_2 + \lambda_2 \mathbf{K}_2 \right) \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{bmatrix} \mathbf{X}_1' \mathbf{V}^{-1} \mathbf{y} \\ \mathbf{X}_2' \mathbf{V}^{-1} \mathbf{y} \end{bmatrix} \quad \text{(A.8)}$$