UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Estimation of Two Popular Econometric Models: Random Effects Panel Data Model and Simultaneous Equations Model

Permalink

https://escholarship.org/uc/item/3jv029cr

Author

Liu, Yue

Publication Date

2013

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Estimation of Two Popular Econometric Models: Random Effects Panel Data Model and Simultaneous Equations Model

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Yue Liu

August 2013

Dissertation Committee:
Professor Aman Ullah, Co-Chairperson
Professor Jun Li, Co-Chairperson
Professor Subir Ghosh
Professor Tae-Hwy Lee

Copyright by
Yue Liu
2013

The Dissertation of Yue Liu is approved:

_____

_____

_____

Committee Co-Chairperson

_____

Committee Co-Chairperson

University of California, Riverside

Acknowledgement

I first would like to express my sincere gratitude to my advisor, Professor Aman Ullah, for all his help on my work. He is the person who inspires me on studies in econometrics. Without his support, it would be impossible for me to collect the master degree in economics. I truly appreciate his valuable questions and advice provided during lectures, meetings and seminars, etc.

The second chapter of this thesis benefits a lot from discussions and cooperation (for the empirical data modeling part) with Professor Madhu S. Mohanty. Without his help, this chapter would not come into existence.

I appreciate all the help from the faculty and staffs at statistics department, economics department and UCR libraries. Also I receive considerable support from my colleagues at statistics department and economics department.

I would like to thank my parents for their patience and consideration. I appreciate their endeavor to educate me as an honest person with full passion.

Last but not least, I feel very grateful to my grandmother. All my current and future achievements are largely attributed to her. Wish her rest in peace in Heaven and I will always love her!

ABSTRACT OF THE DISSERTATION

Estimation of Two Popular Econometric Models:
Random Effects Panel Data Model and Simultaneous Equations Model

by

Yue Liu

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, August 2013
Professor Aman Ullah, Co-Chairperson
Professor Jun Li, Co-Chairperson

In the first chapter of this thesis, we propose a penalized splines (P-splines) estimator for random effects panel data model. While being a nonparametric technique, one of the most attractive properties of splines methods is their analogous setup to parametric regression model. Compared to kernel-based methods, however, splines methods are far less developed at least for econometric models. The asymptotic results of our P-splines estimator are established in this chapter. Monte Carlo simulation is conducted to compare the performance of our P-splines estimator with different kernel-based estimators proposed in recent literature. It turns out that the P-splines estimator consistently performs well and is computationally efficient.

In the second chapter, we develop a general procedure to derive the asymptotic variance-covariance matrices of two-stage estimators that can be used to estimate simultaneous equation systems with a mixture of any number of binary and continuous dependent variables. To demonstrate the usefulness of our procedure, we apply our formulas to empirical data with one continuous and two binary dependent variables in the simultaneous equations system. Our results are expected to be of tremendous help to numerous practitioners of econometrics using two-stage procedures to estimate their simultaneous equations models.

Content

List of Figures ix

List of Tables x

1. Introduction 1

2. A Penalized Splines Estimator for Random Effects Panel Data Model 4

2.1 Introduction……4

2.2 Splines Methods…….7

2.2.1 Regression Splines…….7

2.2.2 Smoothing Splines……16

2.2.3 Penalized Splines……..18

2.3 Panel Data Model…….25

2.4 P-Splines Estimator for Random Effects Panel Data Model……26

2.5 Asymptotic Properties……..27

2.6 Simulation……31

2.6.1 DGP Description……..31

2.6.2 Alternative Estimators…….36

2.6.3 Simulation Results……...38

2.7 Conclusion……...43

3. Asymptotic Variance-Covariance Matrices of Two-Stage Estimators for Simultaneous Equations System 61

3.1 Introduction. ……61

3.2 Two-Stage Estimators for Simultaneous Equations System……65

3.3 Asymptotic Variance-Covariance Matrices for the Two-Stage Estimators….69

3.4 An Empirical Application……74

3.4.1 Model……...75

3.4.2 Data and Results……..78

3.5 Conclusion……..82

4. Conclusion 101

List of Figures

2.1 Linear and quadratic truncated power functions with equally-spaced knots…….10

2.2 Linear and cubic B-spline bases with equally-spaced knots……..13

2.3 Components of a fit with 13 cubic B-splines and a second order penalty to simulated data…….23

2.4 Smoothing and interpolation of simulated data with 23 cubic B-splines and a second order penalty to simulated data……24

2.5 The simulated data and the true functional form for DGP 1……..33

2.6 The simulated data and the true functional form for DGP 2……..34

2.7 The simulated data and the true functional form for DGP 3……..35

2.8 Comparison of P-spline, LS-RE (Su and Ullah, 2007), and FD (Qian and Wang, 2011) estimators…….41

3.1 Relationship of dependent variables in the empirical application…….81

List of Tables

2.1 AMSE of function estimate for random effects panel data model……...…….40

2.2 The estimate of smoothing parameter for each DGP……...……..42

3.1 Two-Stage Estimates of Wage, Happiness and Attitude Equations……83

Chapter 1

Introduction

The very first several statistical or econometric courses that we take always start with linear regression analysis on single equation model with cross-sectional (or independent) data. Those models and analytical techniques are fundamental and in tremendous use especially in industrial systems. However with methodological and computational improvement and increasing data availability, there is a great need to investigate some advanced model or technique to describe the data correctly or more efficiently. This thesis focuses on estimation of two such advanced econometric models: random effects panel data model and simultaneous equations model. It consists of four chapters, including this introduction chapter and a conclusion chapter in the end. In chapter 2 and chapter 3, we shall discuss those two models in detail respectively.

Chapter 2 studies the nonparametric regression estimation for random effects panel data model. Panel data models are currently enjoying increasing popularity

since they combine the cross-sectional data with time series data and therefore would provide more information. The motivation for using nonparametric methods comes from the fact that employing a misspecified parametric model will typically result in inconsistent parameter estimates and hypothesis tests that possess asymptotically incorrect size. If we have no idea of the true functional form, it may be better to use nonparametric techniques. As we know, there are two principle nonparametric estimation methods, the kernel-based estimators and the splines smoothers, where the latter is far less developed at least for econometric models. Therefore we propose a type of splines estimators, P-splines estimator, for random effects panel data model. We discuss different issues in this chapter, such as various bases and splines methods to choose from. We also establish the asymptotic results of our P-splines estimator. Monte Carlo simulation is conducted to compare the finite sample performance of our P-splines estimator with different kernel-based estimators proposed in recent literature. It turns out that the P-splines estimator consistently performs well in terms of both statistical efficiency and computational efficiency.

Chapter 3 considers estimation for simultaneous equations model. Among different candidates, two-stage estimation stands out for its computational and interpretational simplicity. The two-stage procedure to estimate simultaneous equations model has been quite popular among applied econometricians for the last few decades. There are still some issues, however, left to resolve, such as the one that

we shall discuss in this study. When it comes to a simultaneous equations model with more than one binary dependent variable, authors report uncorrected standard errors and test statistics which may lead to unreliable conclusions. Hence, for the benefit of numerous applied economists, chapter 3 develops the asymptotic variance-covariance matrices of two-stage estimators for the simultaneous equations model with a mixture of any number of continuous and binary dependent variables. We also apply our results to empirical data to see how they work. A comparison based on the empirical results with uncorrected standard errors and those with corrected standard errors shows that the difference may be big enough to draw different conclusions.

Chapter 2

A Penalized Splines Estimator for Random Effects Panel Data Model

2.1 Introduction

It has long been felt by both statisticians and econometricians that the investigation of many issues such as density estimation and model fitting should not be constrained by the need to make precise assumptions about functional form or distribution. The motivation for using nonparametric methods comes from the fact that employing a misspecified parametric model will typically result in inconsistent parameter estimates and hypothesis tests that possess asymptotically incorrect size. Nonparametric estimators are consistent under less restrictive assumptions such as smoothness (differentiability) and moment restrictions rather than those required for consistency of parametric estimators.

Therefore, in the absence of knowledge like the true functional form of the conditional mean, nonparametric methods may be preferred.

Nonparametric regression methods for panel data analysis have been a popular statistical research topic since the late 1990s. The needs of panel data analysis from economic research and other scientific areas along with the recognition of the limitation of parametric models in practical data analysis have driven the development of more innovative nonparametric regression methods for panel data model. For example, Su and Ullah (2006), Henderson, Carroll and Li (2008), and Sun, Carroll and Li (2009) consider the problem of nonparametric kernel estimation for fixed effects panel data model. Ullah and Roy (1998), Ruckstuhl, Welsh and Carroll (2000), Lin and Carroll (2000, 2001, 2006), Su and Ullah (2007), Henderson and Ullah (2005, 2008), Martins-Filho and Yao (2009) and Su, Ullah and Wang (2013) consider the nonparametric kernel estimation for random effects panel data model. Some estimators, such as the one proposed by Qian and Wang (2012), can be applied to both fixed effects and random effects panel data model.

All the literature listed above focus on developing kernel-based regression estimators for panel data model. Splines method, as an alternative nonparametric technique, has been far less developed. Green and Silverman (1994), de Boor (2001), Ruppert, Wand and Carroll (2003), Lin, Wang, Welsh and Carroll (2004), Eilers and Marx (1996, 2004), Wu and Zhang (2006), Li and Ruppert (2008), Claeskens,

Krivobokova and Opsomer (2009) discuss different splines methods mainly for cross-sectional data. Therefore in this study we follow the idea of P-splines estimator (Eilers and Marx, 1996) that is originally proposed for independent data and apply it to random effects panel data model. We establish the asymptotic properties for this estimator. Simulation results show that our proposed P-splines estimator has satisfactory finite sample performance compared to the alternative kernel-based estimators proposed in recent literature in terms of both statistical efficiency and computational efficiency.

The rest of this chapter is organized as follows. In section 2.2 we introduce different types of splines methods with different bases. Good properties of our proposed P-splines estimator manifest themselves. Section 2.3 introduces both fixed effects and random effects panel data model. In section 2.4, we propose a P-splines estimator for random effects panel data model. Its asymptotic properties are presented in section 2.5. Section 2.6 reports the simulation results that examine the finite sample performance of our proposed estimator. The final section 2.7 concludes this chapter. The proofs of the main results are collected in Appendix I-A.

2.2 Splines Methods

'Spline' is originally referred to as a mechanical device (wood, plastic or metal) used by architects, engineers and draftsmen to assist in drawing smooth curves passing through some prescribed points (a.k.a knots). The elasticity of the spline material would shape the strip with minimal energy while interpolating those fixed points. The ideas of splines smoothing have their roots in the aircraft and shipbuilding industries.

It was not until 1946 when the concept of splines was first in connection with sufficiently smooth, piecewise polynomial approximation by the mathematician Schoenberg. The splines smoothing is now gaining more popularity among statisticians as a nonparametric regression method. In this section, we will discuss three splines methods: regression splines, smoothing splines and penalized splines. Also we will introduce two types of bases, truncated power basis and B-spline basis, that the spline function may employ.

2.2.1 Regression Splines

An important issue when using local techniques is the determination of the 'local neighborhood'. For local polynomial regression, the neighborhood is specified by a bandwidth h and a point x. For splines smoothing that we shall discuss in this

section, the local neighborhood is specified by a group of locations, say, $\tau_0, \cdots, \tau_{K+1}$, in the range of interest, say, [a, b] where $a = \tau_0 < \tau_1 \cdots < \tau_K < \tau_{K+1} = b$. These locations are called knots, and $\{\tau_k\}_{k=1}^{K}$ are typically referred to as interior knots. These knots divide the interval of interest into subintervals (local neighborhoods), within which a Taylor's expansion up to some degree is valid. In other words, a regression spline is a piecewise polynomial fitted within any two neighboring knots and is joined together at knots properly.

Among various basis candidates, the truncated power basis and B-spline basis are the most popular ones. Hence we shall discuss these two kinds of bases. For notational simplicity, we consider the univariate case that x takes values on [a, b]. Set a sequence of knots as follows

$$\tau_{-p} = \cdots = \tau_0 = a < \tau_1 < \cdots < \tau_K < b = \tau_{K+1} = \cdots = \tau_{K+p+1}$$

where $\{\tau_k\}_{k=1}^{K}$ are equally-spaced interior knots that are the only knots required for truncated power basis. At the same time, for obtaining nonzero function values at $x = a$ and $x = b$, we need an extra $2p+2$ knots to generate a complete B-spline basis. Next we shall first introduce truncated power basis, followed by B-spline basis.

The truncated power basis of degree p, or the $(p+1)$-st order truncated power basis, is defined as

$$\Phi_p(x) = (1, x, \cdots, x^p, (x-\tau_1)_+^p, \cdots, (x-\tau_K)_+^p)', \tag{1.1}$$

where $(x-\tau_k)_+^p = (x-\tau_k)^p I(x-\tau_k \geq 0)$ and we refer to this as the truncated power function of degree p. Figure 2.1, as an example, shows linear and quadratic truncated power functions, with four interior knots equally spaced on [0, 1].

Consider the following nonparametric regression model with the only regressor x:

$$y_i = m(x_i) + u_i, \quad i = 1, \cdots, n \tag{1.2}$$

where m is assumed to be a smooth function and u_i is the error term with mean 0 and variance 1. Using truncated power basis of degree p, $m(x_i)$ can be approximated by the polynomial

$$m(x_i) \approx \sum_{s=0}^{p} \beta_s x_i^s + \sum_{k=1}^{K} \beta_{p+k} (x_i - \tau_k)_+^p = \Phi_p(x_i)'\beta,$$

where $\Phi_p(x_i)$ is defined as in (1.1) and β is a $(K+p+1)\times 1$ vector. Thus model (1.2) can be rewritten in vector form

$$\mathbf{y} \approx \mathbf{X}\beta + \mathbf{u},$$

where $\mathbf{y} = (y_1, \cdots, y_n)'$, $\mathbf{X}$ is an $n \times (K+p+1)$ matrix with each row vector as $\Phi_p(x_i)'$ and $\mathbf{u} = (u_1, \cdots, u_n)'$.

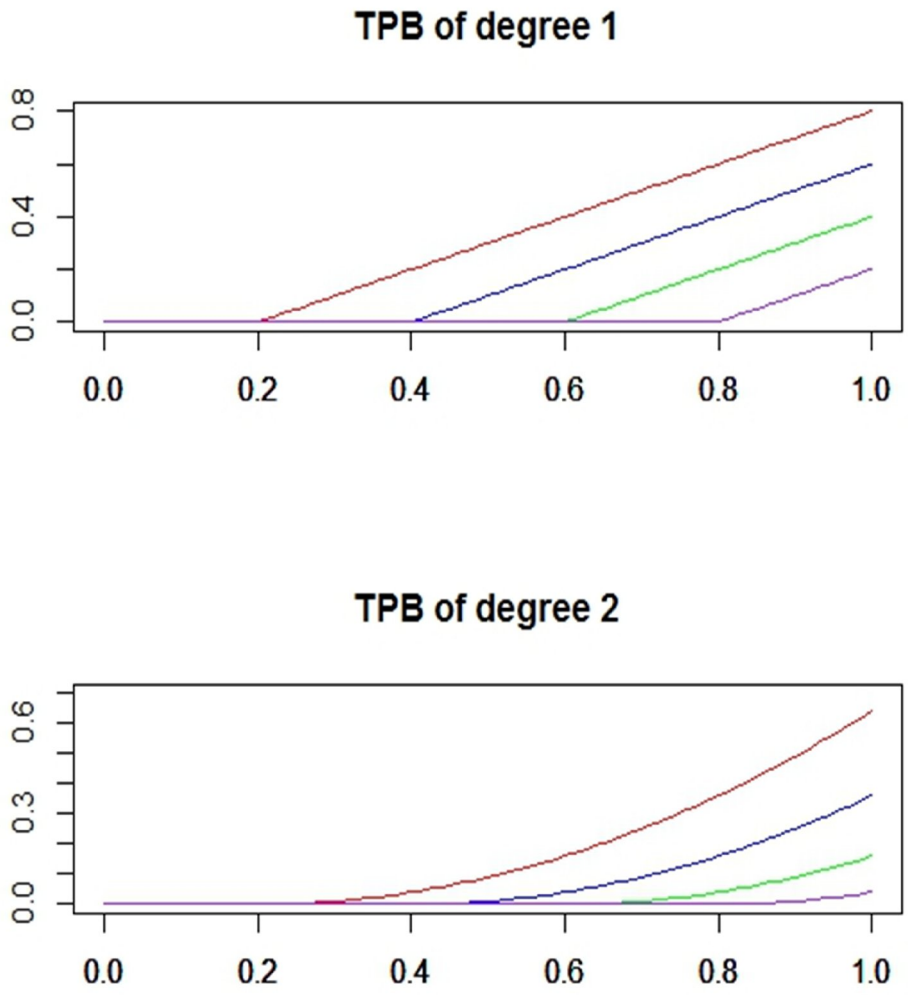


Figure 2.1: Linear (upper panel) and quadratic (lower panel) truncated power functions with equally-spaced knots.

A B-spline basis function can be considered as difference of some truncated power functions. In general the difference is performed on $K+2p+2$ different $(p+1)$-tuples of degree p truncated power functions resulting in $K+p+1$ different $(p+1)$-st order B-spline bases:

$$N_{p+1}(x)=\{N_{k,p+1}(x):\ -p\le k\le K\}' \tag{1.3}$$

where

$$N_{k,p+1}(x)=(\tau_{k+p+1}-\tau_k)[\tau_k,\cdots,\tau_{k+p+1}](t-x)_+^p,\qquad k=-p,\cdots,K$$

where $[\tau_k,\cdots,\tau_{k+p+1}]g$ denotes the $(p+1)$-st order divided difference of the function g (de Boor, 2001).

In particular, if the knots are equidistant, the formula can be further simplified as

$$N_{k,p+1}(x)=\frac{1}{\delta^p p!}\sum_{i=0}^{p+1}(-1)^i\binom{p+1}{i}(x-\tau_{k+i})_+^p,\quad -p\le k\le K$$

with δ being the length of each subinterval of [a, b].

As an alternative, B-splines can be defined in a recursive way,

$$N_{k,1}(x) = \begin{cases} 1, & \tau_k \le x < \tau_{k+1} \\ 0, & otherwise \end{cases}$$

$$N_{k,p+1}(x) = \frac{x-\tau_k}{\tau_{k+p}-\tau_k} N_{k,p}(x) + \frac{\tau_{k+p+1}-x}{\tau_{k+p+1}-\tau_{k+1}} N_{k+1,p}(x), \quad k = -p, \cdots, K .$$

If the knots are equally-spaced, then the formula would be

$$N_{k,p+1}(x) = \left[(x-\tau_k) N_{k,p}(x) + (\tau_{k+p+1}-x) N_{k+1,p}(x) \right] \Big/ p\delta , \quad k = -p, \cdots, K ,$$

where δ is the length of each subinterval of [a, b].

This recursive definition is actually more often used in the B-splines literature. One reason is that the derivative formula for B-spline functions can easily be obtained using such definition and it is necessary for deriving the asymptotic properties. We hold the derivative formula till section 2.2.3.

Figure 2.2 on the next page shows linear and cubic B-spline bases functions, with four interior knots equally spaced on [0, 1].

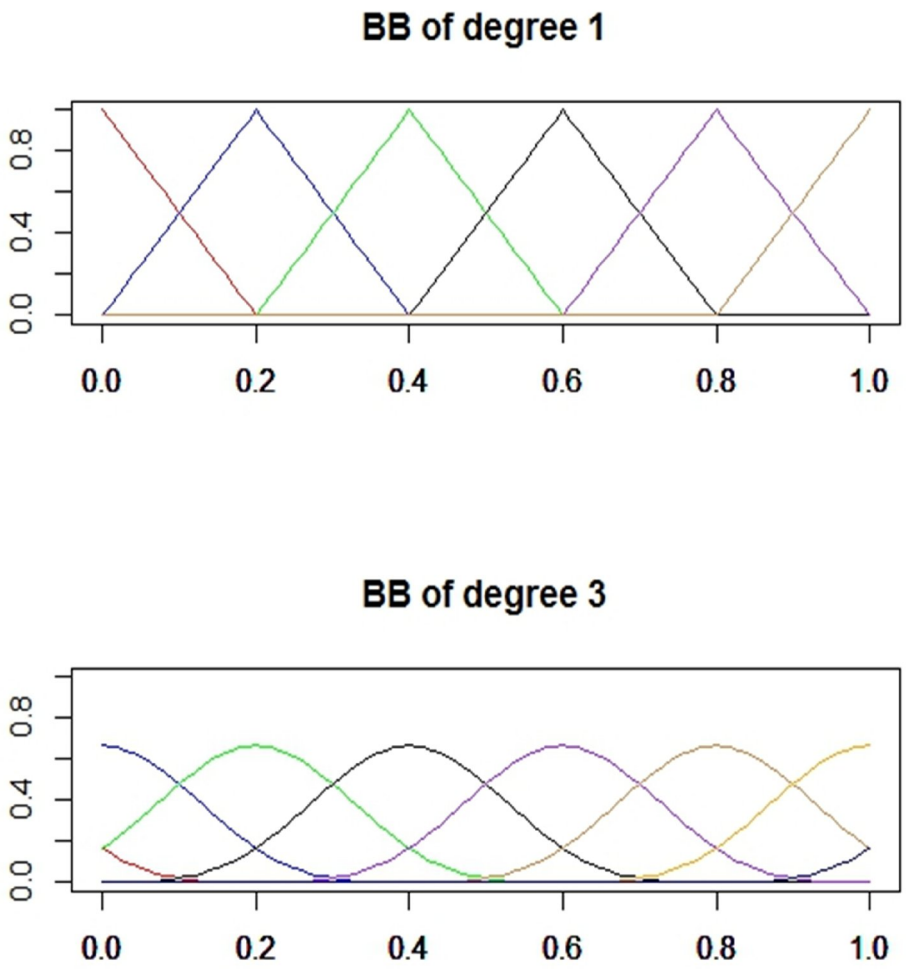


Figure 2.2: Linear (upper panel) and cubic (lower panel) B-spline bases with equally-spaced knots.

No matter how the B-spline bases are constructed, we can always use them to estimate the regression function $m(x)$:

$$\mathbf{y} \approx \mathbf{N}_{p+1}\beta + \mathbf{u}$$

where β is a $(K+p+1)\times 1$ vector and $\mathbf{N}_{p+1} = \left[\{N_{p+1}(x_1),\cdots,N_{p+1}(x_n)\}'\right]_{n\times(K+p+1)}$ with $N_{p+1}(x_i)$ defined in (1.3). For notational simplicity, we suppress the subscript $p+1$ for the B-spline matrix and write

$$\mathbf{y} \approx \mathbf{N}\beta + \mathbf{u} \tag{1.4}$$

from now on.

After specifying the basis functional form, it is very similar to a linear regression model and can be estimated using ordinary least squares. Here we assume the interior knots are equally-spaced. Ruppert, Wand and Carroll (2003) suggested that unequally-spaced knots perform better. It is good advice for regression splines method. By using P-splines method that we shall discuss later, however, equally-spaced knots are good enough (Eilers and Marx, 2004). In addition, equally-spaced knots are easy to specify and report, and they simplify the computations.

There is still one more question left, that is, which one is better, truncated power basis or B-spline basis. In fact, the latter has a wider usage in nonparametric estimation. There are mainly three advantages of B-spline basis over truncated power basis. First, the B-spline base functions are bounded while truncated power functions may be not. Therefore, we will not worry too much about the convergence when B-spline basis is employed, see Chui (1992). Second, the results of estimation based on power series can be sensitive to the outliers. Finally, B-spline basis has small and compact support that is most efficient for both software and hardware implementation. In light of all its advantages, we would apply B-spline basis for all the estimation in the following study.

Consider the regression splines estimator $\hat{m}_{reg}(\cdot)$ first:

$$\hat{\mathbf{m}}_{reg} = \arg \min_{s(\cdot)\in S(p+1;\tau)} (\mathbf{y}-\mathbf{s})'(\mathbf{y}-\mathbf{s}),$$

where

$$S(p+1;\tau) = \left\{ s(\cdot) \in C^{p-1}[a,b] : s \text{ is a degree } p \text{ polynomial on each } [\tau_k, \tau_{k+1}] \right\}$$

is the set of spline functions of degree p with knots τ's (Schumaker, 2007).

With B-spline basis, we may estimate the parameters for model (1.4)

$$\hat{\beta} = \arg\min_{\beta} (\mathbf{y}-\mathbf{N}\beta)'(\mathbf{y}-\mathbf{N}\beta) .$$

It follows that the fitted regression spline is

$$\hat{m}_{reg}(x) = N'(x)\hat{\beta} = N'(x)(\mathbf{N}'\mathbf{N})^{-1}\mathbf{N}'\mathbf{y} \in S(p+1;\tau). \tag{1.5}$$

2.2.2 Smoothing Splines

The difficulty of applying regression splines is that its good performance crucially depends on the location and number of knots, which may not be easy to determine. Once the knots are selected improperly, over- (or under-)fitting problems will inevitably arise. Smoothing splines, as an alternative, overcome this difficulty via taking all the distinct observing points as knots, and using a roughness penalty to control the smoothness of the resulting smoothers.

The idea of smoothing splines is to find the function $m(x)$ such that the following minimization problem is solved

$$\min_{m(\cdot)}\left(\sum_{i=1}^{n}(y_i - m(x_i))^2 + \lambda\int\left[m^{(q)}(t)\right]^2 dt\right),$$

or written compactly,

$$\min_{\beta}\left(\left\|\mathbf{y} - \mathbf{N}\beta\right\|^2 + \lambda\beta'\mathbf{D}_q\beta\right),$$

where $\mathbf{N}$ is a spline matrix, β is the spline coefficient, $\mathbf{D}_q$ is the penalty matrix,

and $\lambda > 0$ is the smoothing parameter that controls the trade-off between model complexity and goodness-of-fit. $\lambda \to 0$ implies an unpenalized estimate, and $\lambda \to \infty$ implies a limiting estimate of q th order polynomial. The smoothing parameter is commonly selected by methods such as cross-validation, generalized cross-validation or Akaike information criterion. For further details, see Green and Silverman (1994).

The solution will therefore be

$$\hat{\beta} = \left(\mathbf{N}'\mathbf{N} + \lambda \mathbf{D}_q \right)^{-1} \mathbf{N}'\mathbf{y},$$

and the fitted smoothing spline is given by

$$\hat{m}(x) = \sum_{k=1}^{n} b_k(x)\hat{\beta}_k.$$

Again note that the number of splines equals to the sample size n. It is obvious that the quantity of data may be so large that the system of linear equations for the exact solution is too expensive to solve, and a cheaper approximation may be preferred. Computational expense is not likely to be a problem with some computing algorithms such as the Reinsch algorithm, but in very large systems it can become numerically unstable.

2.2.3 Penalized Splines

A Penalized spline is designated to overcome the drawbacks mentioned in last section by using the basis functions of some degree with a pre-specified number of knots that is much smaller than the total number of observing points.

Penalized splines can be considered a modification to smoothing splines defined as the minimizer of

$$\sum_{i=1}^{n}\left(y_i - \sum_{k=-p}^{K} \beta_k N_{k,p+1}(x_i) \right)^2 + \lambda \int_a^b \left\{ \left(\sum_{k=-p}^{K} \beta_k N_{k,p+1}(x) \right)^{(q)} \right\}^2 dx . \qquad (1.6)$$

As noted in section 2.2.1, the derivative formula for B-spline functions can be obtained via recursive definition and it is given in de Boor (2001) as

$$\left(\sum_{k=-p}^{K} \beta_k N_{k,p+1}(x) \right)^{(q)} = \sum_{k=-p+q}^{K} N_{k,p+1-q}(x) \beta_k^{(q)}, \qquad (1.7)$$

where $\beta_k^{(q)}$ are defined recursively as

$$\beta_k^{(1)} = p(\beta_k - \beta_{k-1}) / (\tau_{k+p} - \tau_k),$$

$$\beta_k^{(q)} = (p+1-q)(\beta_k^{(q-1)} - \beta_{k-1}^{(q-1)}) / (\tau_{k+p+1-q} - \tau_k), \ q > 1 .$$

For equidistant knots, $\beta_k^{(q)}$ can be written as

$$\beta_k^{(1)} = (\beta_k - \beta_{k-1}) / \delta,$$

$$\beta_k^{(q)} = (\beta_k^{(q-1)} - \beta_{k-1}^{(q-1)}) / \delta, \ q > 1.$$

With equation (1.7), the objective function (1.6) can be put more neatly,

$$\left\| \mathbf{y} - \mathbf{N}\beta \right\|^2 + \lambda \beta' \mathbf{D}_q \beta .$$

The solution can therefore be written as

$$\hat{\beta} = \left(\mathbf{N}'\mathbf{N} + \lambda \mathbf{D}_q \right)^{-1} \mathbf{N}'\mathbf{y},$$

and the fitted penalized spline is given by

$$\begin{aligned} \hat{m}(x) &= N'(x)(\mathbf{N}'\mathbf{N} + \lambda \mathbf{D}_q)^{-1} \mathbf{N}'\mathbf{y} \\ &= \sum_{k=-p}^{K} N_k(x) \hat{\beta}_k \end{aligned} .$$

Note that $K << n$.

The smoothing parameter λ is defined the same as usual. q denotes the order of penalty which typically satisfies $q \le p+1$, where p is the degree of B-spline basis. When $q = p+1$, the integral reduces to the difference penalty. If $q \le p$, the difference penalty is still a good discrete approximation to the integral. Consequently, it is reasonable to avoid direct calculation of the roughness matrix by

specifying a difference penalty without much loss. Eilers and Marx therefore proposed in their 1996 paper the P-splines estimator, a combination of B-splines and difference penalties, for cross-sectional data model. We follow this line and propose a P-splines estimator for random effects panel data model in this study. Note that the P-splines estimator is one of various penalized splines estimators. We will use P-splines estimator rather than the general penalized splines estimator from now on.

We have briefly mentioned that P-splines estimator may dramatically reduce the computational burden compared to smoothing splines. Another benefit of P-splines estimator is that location and number of knots do not play a significant role in curve fitting. For regression splines, equally-spaced knots may produce empty regions in the domain of x and unsatisfactory results would follow. P-splines estimator, however, can obtain desirable results with equally-spaced knots due to the power of the penalty. And as discussed in section 2.2.1, assuming equally-spaced knots would simplify the expression of B-spline basis functions to a large extent and would therefore ease the estimation and computation. In addition, P-splines estimator is not sensitive to the number of knots which would also make our work less troublesome.

To close this section, we present two small sets of Monte Carlo simulations in order to get a better understanding of how P-splines estimator works,.

Simulation 1:

We generate data of size 50 from the following data generating process (DGP)

$$Y = 0.6 + (X - 0.6)^2 + \varepsilon ,$$

where X follows $U(0,1)$ and ε is distributed as $N(0,(.1)^2)$.

Figure 2.3 shows the components of a fit given by P-splines estimator with cubic B-splines and second order penalty to the simulated data (shown as squares). We place 9 equidistant interior knots that produce 13 cubic B-splines. The individual B-splines, scaled by their coefficients, are shown on the bottom of the plot. The asterisks give the B-spline coefficients, which are very close to the fitted curve and exhibit the skeleton of the fit. In fact it is the difference penalty that forces the coefficients to follow a smooth pattern along the fitted curve. The degree of B-splines then determines the smoothness of the curve. This is the attractive property of P-splines estimator in terms of visualization (Eilers and Marx, 2004).

Simulation 2:

We generate data of size 100 from the following DGP

$$Y = \sin(2\pi X) + \varepsilon,$$

where

$$X \sim \begin{cases} N(.25,(.07)^2) & w.p.\ \ 0.5 \\ N(.75,(.07)^2) & w.p.\ \ 0.5 \end{cases}, \text{ and } \ \varepsilon \sim N(0,(.05)^2).$$

Note that the independent variable X is generated as clustered data. Hence it is inevitable to produce empty subintervals in the domain of X that contain no data at all if we use equally-spaced knots. The regression splines estimator, in this case, would suffer the empty corner issue and deliver some bad result. The P-splines estimator, however, allows reasonable smooth interpolation and extrapolation. Figure 2.4 illustrates the smoothing and interpolation of the simulated data (shown as dots). In fact, it has been proved that the order of difference penalty, determines how smooth the interpolation and extrapolation would be. The interpolation is of degree $2q-1$ and the extrapolation is of degree $q-1$ if the order of difference penalty is q. Often, we want to make the extrapolation at least linear which indicates a second order difference penalty, i.e. $q=2$. It follows that the interpolation would be a cubic curve. To keep the whole curve cubic, we have to employ cubic splines. That also partially explains why cubic B-splines are so popular.

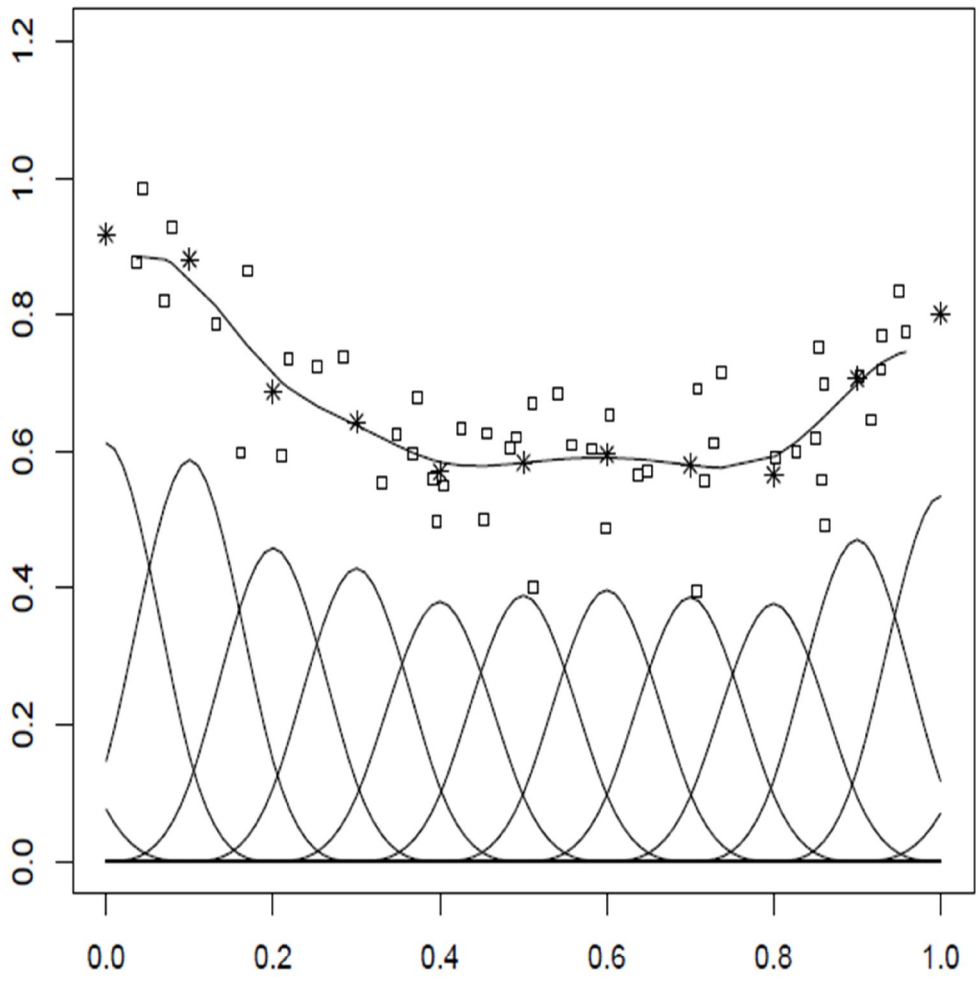


Figure 2.3: Components of a fit with 13 cubic B-splines and a second order penalty to simulated data (squares). The asterisks show the coefficients of the B-splines. The scaled B-splines are shown on the bottom.

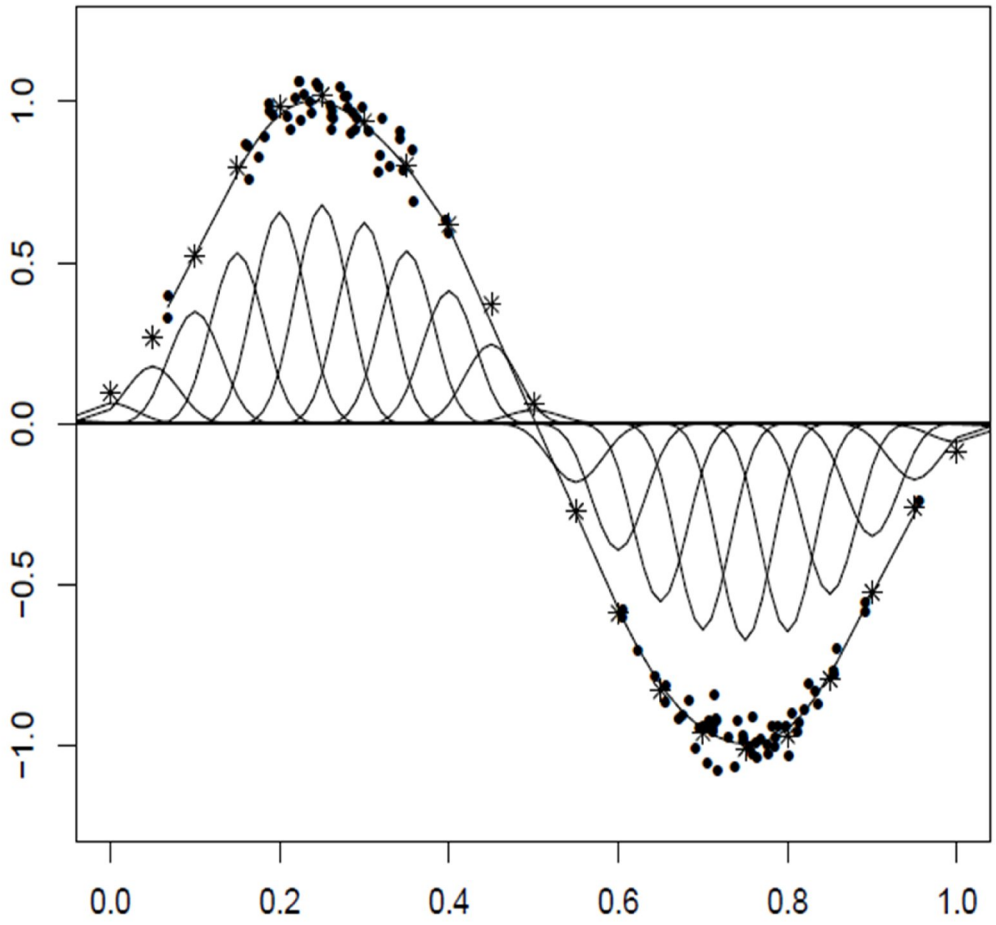


Figure 2.4: Smoothing and interpolation of simulated data with 23 cubic B-splines and a second order penalty to simulated data (dots). The asterisks show the coefficients of the B-splines. The scaled B-splines are also shown.

2.3 Panel Data Model

Data sets that combine time series and cross sections are common in economics. For example, the published statistics from the U.S. Census Bureau contain numerous series of economic aggregates observed yearly for different states. Such data sets are referred to as panel data or longitudinal data.

By way of preface to the analysis to follow, we note the two variants of panel data models, fixed effects panel data model and random effects panel data model. In general, these two panel data models can be written as

$$y_{it} = m(x_{it}) + \alpha_i + u_{it},\ i = 1, \cdots, n,\ t = 1, \cdots, T$$

where m is an unknown smooth function, α_i's are the unobserved individual effects, and u_{it} is the idiosyncratic error term. If α_i's are uncorrelated with x_{it}, then random effects panel data model is employed. Otherwise, fixed effects panel data model is appropriate. It is worth mentioning that the term 'fixed' is not nonstochastic. Rather α_i's should be thought of as random error terms that are correlated with x_{it} with unknown correlation structure. Here we develop a penalized splines estimator for random effects panel data model without ruling out the possibility of extension to fixed effects panel data model.

For the sake of simplicity, we assume balanced panel and strict exogeneity for the regressor all through this study.

2.4 P-Splines Estimator for Random Effects Panel Data Model

Consider the following random effects panel data model

$$y_{it} = m(x_{it}) + \alpha_i + u_{it} \tag{1.8}$$

where x_{it} is an exogenous variable, α_i is i.i.d $(0, \sigma_\alpha^2)$ that is uncorrelated with x_{it}, and u_{it} is i.i.d $(0, \sigma_u^2)$.

Let $\grave{o}_{it} = \alpha_i + u_{it}$ and $\grave{o}_i = (\grave{o}_{i1}, ..., \grave{o}_{iT})'$ such that

$$\Sigma = E(\grave{o}_i \grave{o}_i') = \sigma_\alpha^2 1_T 1_{T'} + \sigma_u^2 I_T \tag{1.9}$$

where 1_T is a $T \times 1$ vector of ones. It follows that $E(\grave{o}\grave{o}') = \Omega = I_n \otimes \Sigma$.

In fact, when establishing the asymptotic properties in the following section, the variance-covariance matrix Σ can be any positive definite matrix rather than be restricted to the form given in (1.9).

To get the estimates of parameters for this random effects panel data model (1.8), the objective function can be written as

$$(\mathbf{y}-\mathbf{N}\beta)'\Omega^{-1}(\mathbf{y}-\mathbf{N}\beta)+\lambda\beta'\mathbf{D}_q\beta, \tag{1.10}$$

if we apply the P-splines estimator with B-spline basis.

The solution of coefficients can then be obtained

$$\hat{\beta}=(\mathbf{N}'\Omega^{-1}\mathbf{N}+\lambda\mathbf{D}_q)^{-1}\mathbf{N}'\Omega^{-1}\mathbf{y}. \tag{1.11}$$

Thus the fitted curve given by P-splines estimator for model (1.8) is

$$\hat{\mathbf{m}}=\mathbf{N}(\mathbf{N}'\Omega^{-1}\mathbf{N}+\lambda\mathbf{D}_q)^{-1}\mathbf{N}'\Omega^{-1}\mathbf{y}. \tag{1.12}$$

2.5 Asymptotic Properties

This section renders the asymptotic theorems (in the sense that $n\to\infty$ but T fixed) for our P-splines estimator discussed above for random effects panel data model. At first glance, the proof may seem a straightforward extension of that for cross-sectional data model. On a closer scrutiny, however, it reveals a different layer of complexity. For example, the asymptotic results for cross-sectional data model rely heavily on the band structure of a B-spline design matrix, which is definitely not

applicable for random effects panel data model. Therefore a rigorous mathematical proof is necessary to obtain tenable results against panel data model backdrop. We present the principle assumptions and theorems in this section with all the proof details provided in Appendix I-A.

The following assumptions are stated in order to study the asymptotic properties of P-splines estimator.

(A1) (Quasi-Uniform, see Zhou, Shen and Wolfe, 1998)

$$\max_{1\le k\le K}|\,\delta_{k+1}-\delta_k\,|=o(K^{-1}) \quad \text{and} \quad \exists M>0,\ s.t.\ \ \delta\Big/\min_{1\le k\le K}\delta_k\le M\ ,$$

where $\delta_k=\tau_k-\tau_{k-1},\ \ \delta=\max_{1\le k\le K}\delta_k$.

Note that assumption (A1) assures that $\delta\sim K^{-1}$.

(A2) x_{it} has a continuous density function $\rho^{[t]}(\cdot)$ that is bounded away from zero and infinity on [a, b] for $t=1,\cdots,T$. Assume that $K=o(n^{1/2})$. It follows from the Glivenko-Cantelli theorem and (A1) that

$$\sup_{x\in[a,b]}\|\,Q_n^{[t]}(x)-Q^{[t]}(x)\,\|=o_p(\delta),\quad t=1,\cdots,T$$

where $Q_n^{[t]}(x)$ is the empirical distribution function of $\{x_{it}\}_{i=1}^n$, and $Q^{[t]}(x)$ is the cumulative distribution function of x_{it}.

(A3) Eigenvalues of $\lambda(nT)^{-1}\mathbf{G}_{K,n}^{-1}\mathbf{D}_q$ are less than 1, where $\mathbf{G}_{K,n}=\frac{\mathbf{N}'\Omega^{-1}\mathbf{N}}{nT}$.

Following the idea of Claeskens (2009), we decompose the average mean squared error (AMSE) into three terms, each with its meaning. To be specific, they are average variance, average squared shrinkage bias and average squared approximation bias, respectively. The order of AMSE can be obtained via deriving the order of each term. The expression of asymptotic bias and variance of the proposed estimator follows with the optimal convergence rate being achieved. The asymptotic normality is also established. With assumptions (A1)-(A3) and those lemmas given in Appendix I-A, we state the following asymptotic theorems.

Theorem 1 Under assumptions (A1)-(A2), and for $p \le 2q-1$ and $K_q < 1$,

$$\text{AMSE} = O_p(\frac{K}{n}) + O_p(\frac{\lambda^2 K^{2q}}{n^2}) + O_p(K^{-2(p+1)})$$

where K_q is defined in the Appendix. From this decomposition of AMSE, we can obtain that $K = O_p(n^{1/(2p+3)})$, and $\lambda = O_p(n^j)$ where $j \le \frac{(p+2-q)}{(2p+3)}$.

Theorem 2 Under assumptions (A1)-(A3), and for $m(\cdot) \in C^{p+1}[a, b]$, $p \le 2q-1$ and $K_q < 1$, the asymptotic bias and variance of the penalized splines estimator (1.12) are given as

$$E\left(\hat{m}(x) \mid \mathbf{x}\right) - m(x) = b_r(x) + b_\lambda(x) + o_p(\delta^{p+1}) + o_p(\lambda n^{-1}\delta^{-q})$$

$$Var\left(\hat{m}(x) \mid \mathbf{x}\right) = \frac{1}{nT} N'(x)\mathbf{G}^{-1}N(x) - \frac{\lambda c}{(nT)^2} N'(x)\mathbf{G}^{-1}\mathbf{D}_q\mathbf{G}^{-1}N(x) + o_p((n\delta)^{-1}) + o_p(\lambda n^{-2}\delta^{-(2q+1)})$$

where $b_r(x) = -\frac{m^{(p+1)}(x)\delta_i^{p+1}}{(p+1)!} B_{p+1}\left(\frac{x-\tau_i}{\delta_i}\right)$ with $B_{p+1}(\cdot)$ defined as the $(p+1)$ st Bernoulli polynomial (see Barrow and Smith, 1978), $b_\lambda(x) = \frac{\lambda c}{nT} N'(x)\mathbf{G}^{-1}\mathbf{D}_q\beta$ with $\mathbf{G}$ an $nT \times nT$ matrix defined in the Appendix , and $c > 0$ is a constant.

From Theorem 1 and Theorem 2, we have

$$E\left(\hat{m}(x) \mid \mathbf{x}\right) - m(x) = O_p(\delta^{p+1}) + O_p(\lambda n^{-1}\delta^{-q}),$$

$$Var\left(\hat{m}(x) \mid \mathbf{x}\right) = O_p((n\delta)^{-1}) + O_p(\lambda n^{-2}\delta^{-(2q+1)}).$$

With the order of K and λ given in Theorem 1, it is easy to verify that the optimal convergence rate (see Stone, 1982) is achieved as follows

$$E\left(\hat{m}(x)\,|\,\mathbf{x}\right)-m(x)=O_p(n^{-(p+1)/(2p+3)})$$

$$Var\left(\hat{m}(x)\,|\,\mathbf{x}\right)=O_p(n^{-2(p+1)/(2p+3)}).$$

Theorem 3 Consider the random effects panel data model (1.8), given Theorem 2 and all assumptions listed above,

$$\frac{\hat{m}(x)-(m(x)+b(x))}{\sqrt{Var(\hat{m}(x)\,|\,\mathbf{x})}}\xrightarrow{d}N(0,1)$$

where $b(x)=b_r(x)+b_\lambda(x)$.

2.6 Simulation

In this section, we conduct Monte Carlo simulation to study the properties of our proposed P-splines estimator. Two popular kernel-based estimators[1] proposed in recent literature are included to compare with our P-splines estimator.

2.6.1 DGP Description

We consider three classical DGPs for panel data model

DGP 1: $y_{it}=\sin(2x_{it})+\alpha_i+u_{it}$

DGP 2: $y_{it}=\sin(2\pi x_{it})+\alpha_i+u_{it}$

DGP 3: $y_{it}=x_{it}+2e^{16x_{it}^2}+\alpha_i+u_{it}$

where x_{it} is i.i.d. $U(-1,1)$ and u_{it} is i.i.d. $N(0,1)$, $i = 1,...,n$, $t = 1,...,T$. In general, we assume that the individual effects and the independent variable are related as $\alpha_i = \mu_i + cx_i$, where μ_i is i.i.d. $U(-1,1)$, $x_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$ is the individual mean. It is clear that any nonzero value of c would give the fixed effects panel data model. $c = 0$ would lead to the random effects panel data model which is the case in our study.

Figure 2.5-2.7 show DGP 1-3 respectively. Apparently, these three DGPs have different characteristics. DGP 1 has a very smooth functional form. DGP 2 has several peaks and troughs. DGP 3 has a bump in the middle but rather flat tails. Using these DGPs, we can examine how our P-splines estimator works towards different curve fittings.

Due to these differences, we may expect the estimate of smoothing parameter for DGP 1 to be relatively large, for DGP 2 to be small, and for DGP 3 would be in between. In section 1.6.3, we report the average of estimates of smoothing parameter for each DGP. It turns out that the estimates indeed match our expectation.

We take 100 replications for each simulation. Three observations are obtained from each individual, i.e. $T = 3$. The number of individuals n may take values of 50, 100 and 200 for each DGP.

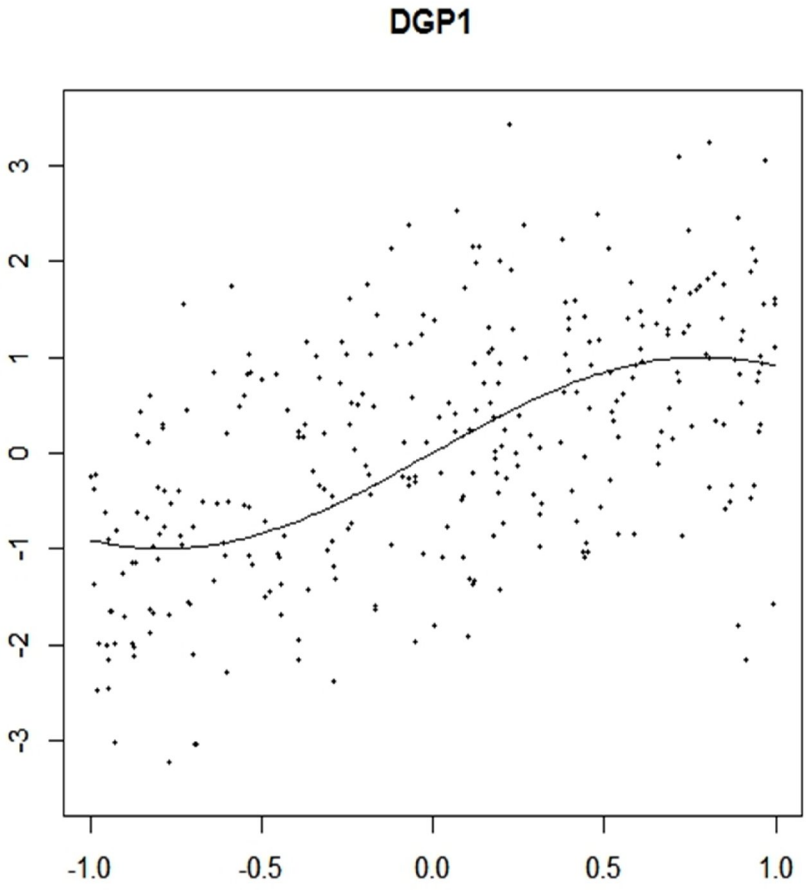


Figure 2.5: The simulated data and the true functional form for DGP 1

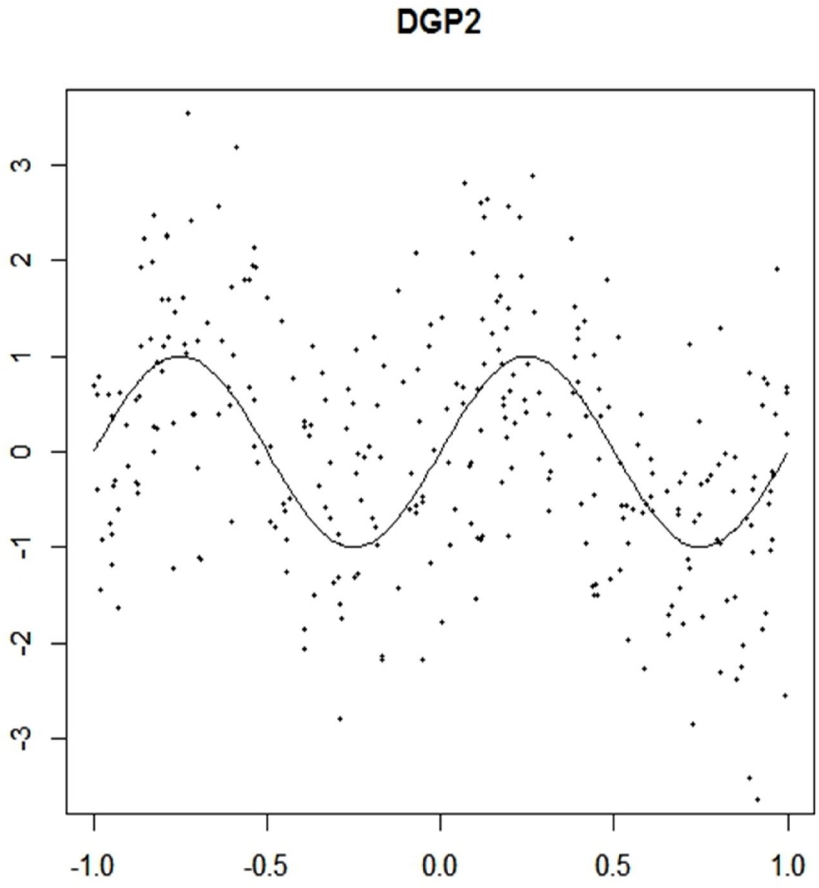


Figure 2.6: The simulated data and the true functional form for DGP 2

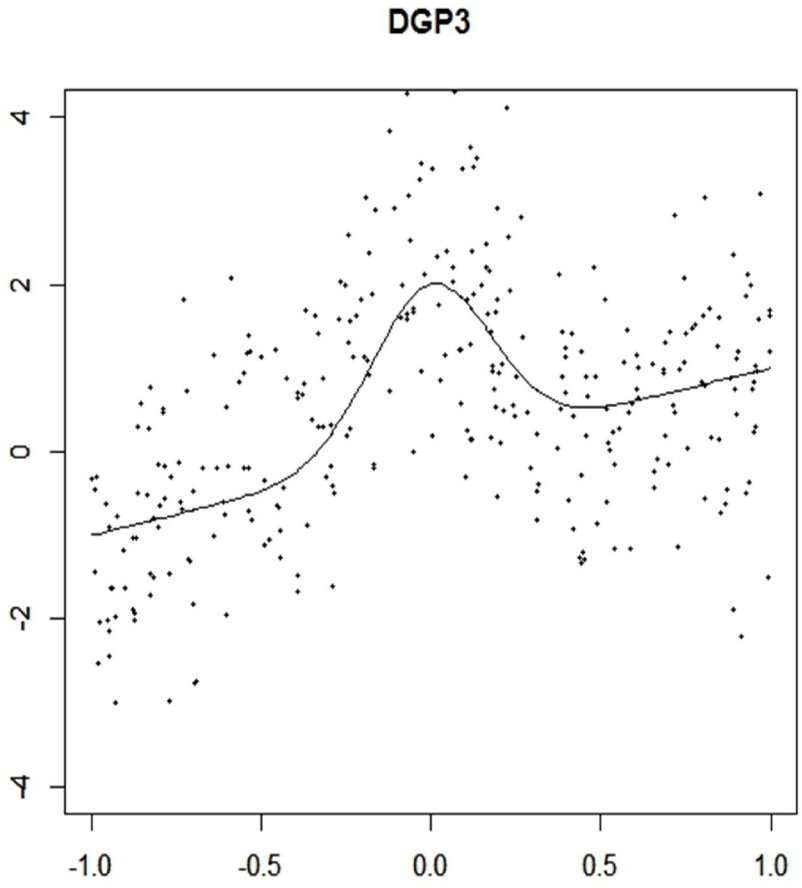


Figure 2.7: The simulated data and the true functional form for DGP 3

2.6.2 Alternative Estimators

We briefly introduce two alternative estimators in this section for the comparison. One is the LS-RE estimator proposed by Su and Ullah in their 2007 paper, and the other is the FD estimator proposed by Qian and Wang in 2012.

- ***LS-RE estimator***

A two-step kernel-based estimator is developed for random effects panel data model. Stack model (1.8) in the vector form

$$y = m(\underline{x}) + \grave{o}.$$

It follows that

$$\begin{aligned}\Omega^{-1/2} y &= \Omega^{1/2} m(\underline{x}) + \Omega^{-1/2}\grave{o} \\ &= \Omega^{1/2} m(\underline{x}) - m(\underline{x}) + m(\underline{x}) + \Omega^{-1/2}\grave{o}\end{aligned}$$

Or,

$$y^* = m(\underline{x}) + \Omega^{-1/2}\grave{o},$$

where $y^* = \Omega^{-1/2} y + (I - \Omega^{-1/2}) m(\underline{x})$ is the transformed variable and $\Omega^{-1/2}\grave{o}$ has an identity variance matrix. However, y^* is not observed. So, at the first step do the local linear least squares (LLLS) estimation of y on x to get the residuals, $\hat{m}(x)$

and $\hat{\Omega}$. $\hat{\Omega}$ is actually obtained by estimating σ_α^2 and σ_u^2 which is discussed in detail in Li and Racine (2007). This gives $\hat{y}^*$. Then, at the second step obtain the LLLS estimation by regressing $\hat{y}^*$ on x.

- ***FD estimator***

The FD estimator is initially proposed for semiparametric panel data model. But it is easy to apply to nonparametric panel data model as well. The essential idea is that we estimate a new regression function after taking the first differencing (FD) on model (1.8), and then recover the original regression function $m(\cdot)$ by marginal integration.

After applying the first differencing approach, we get

$$y_{it}^* = m^*(x_{it}, x_{it\text{-}1}) + u_{it}^* \tag{1.13}$$

where $y_{it}^* \equiv y_{it} - y_{it\text{-}1}$, $m^*(x_{it}, x_{it\text{-}1}) \equiv m(x_{it}) - m(x_{it\text{-}1})$, and $u_{it}^* \equiv u_{it} - u_{it\text{-}1}$. The new regression function $m^*(\cdot)$ can then be estimated using the kernel-based estimator.

Via marginal integration, we have

$$\hat{m}(v) = \int \hat{m}^*(v, w) q(w) dw, \tag{1.14}$$

where $\hat{m}^*(v,w)$ is the regular kernel estimator of $m^*(v,w)$ and $q(\cdot)$ is a predetermined density function.

Often, to obtain the estimate of $m(\cdot)$, the sample version of equation (1.14) is used instead

$$\hat{m}_s(v)=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{m}^*(v,x_{it}).$$

The FD estimator proposed by Qian and Sun has some drawbacks. One is that it worsens the 'curse of dimensionality' by doubling the number of regressors in the nonparametric model. The other two drawbacks will be discussed in next section along with the simulation results.

2.6.3 Simulation Results

Table 2.1 lists the statistical efficiency results for the three estimators in comparison for DGP 1-3. The performance is evaluated in terms of average mean squared error (AMSE). The findings in Table 1 are summarized as follows. First, our P-splines estimator with cubic B-spline basis performs better than the alternatives for all three DGPs, although the LS-RE estimator is almost as efficient. Second, the FD estimator seems quite sensitive to the true functional form while the other two estimators are much more robust. In particular, the FD estimator does not provide a

desirable result for DGP 2. The FD estimator underutilizes the information available which may lead to this inefficiency and needs further improvement. First, the FD estimator is proposed for identity variance-covariance matrix and therefore definitely loses efficiency to some extent for random effects panel data model. In addition, the FD estimator treats $m^*(\cdot)$ in model (1.13) as a fully unknown smooth function, which does not use the information that $m^*(\cdot)$ is actually the difference of the same functional form $m(\cdot)$ with different x values.

In Figure 2.8, we compare our P-splines estimator (leftmost column), LS-RE estimator (middle) and FD estimator (rightmost) using three DGPs. In each diagram, the black line is the true function and the three other lines are estimates from three random replications. Like the findings in Table 2.1, Figure 2.8 shows that P-splines estimator performs consistently well for all DGPs. LS-RE estimator also works fine. But FD estimator is not very efficient especially for estimating peaks and troughs, which again confirms the two drawbacks as noted above.

Table 2.1: AMSE of function estimate for random effects panel data model

DGP	n	LS-RE	FD	P-splines (p=3)
1	50	.0411	.0756	.0376
	100	.0264	.0370	.0206
	200	.0149	.0207	.0132
2	50	.0928	.2256	.0731
	100	.0489	.1209	.0386
	200	.0275	.0696	.0201
3	50	.0745	.0717	.0677
	100	.0453	.0599	.0400
	200	.0268	.0248	.0219

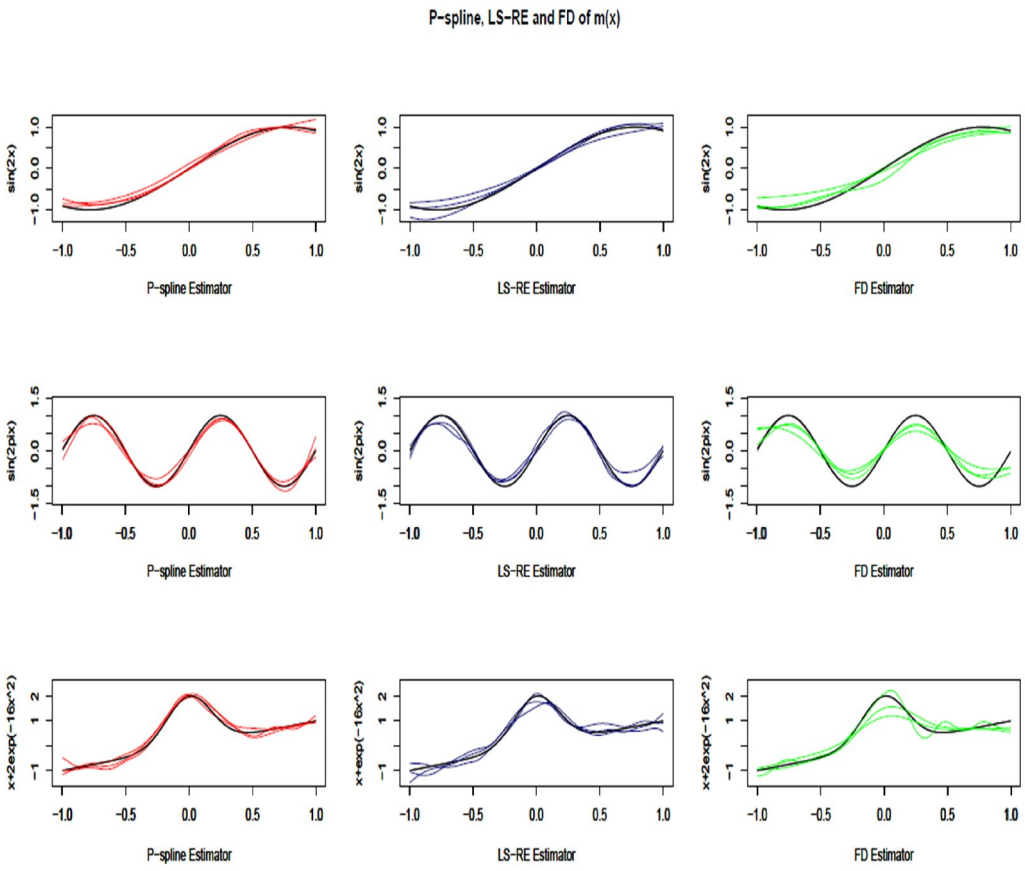


Figure 2.8: Comparison of P-spline (red), LS-RE (Su and Ullah, 2007) (blue), and FD (Qian and Wang, 2012) (green) estimators. In each diagram, the black line is the true function and the three other lines are estimates from three random replications.

We have discussed the statistical efficiency of these three estimators. When it comes to computational efficiency, P-splines estimator and LS-RE estimator are competitive and more efficient than FD estimator.

As mentioned at the end of section 2.6.1, we also report in Table 2.2 the estimates of smoothing parameters for three DGPs when $n = 200$. The estimates match our expectation according to smoothness degree of different true regression functional forms.

Table 2.2: The estimate of smoothing parameter for each DGP

$n = 200$	**DGP 1**	**DGP 2**	**DGP 3**
λ	2100.21	63.31	92.34

2.7 Conclusion

In this chapter, we propose P-splines estimator for random effects panel data model. We illustrate its advantages over other splines methods. Its asymptotic properties are established. Simulation results suggest our proposed estimator outperforms the alternative kernel-based estimators in the mean squared sense. It also turns out that our proposed estimator is very competitive in terms of computational efficiency.

There are many directions for future work based on this study. For example, as mentioned at the beginning of this chapter, the P-splines estimator may in all likelihood be applied to fixed effects panel data model. In fact, some preliminary simulation results for fixed effects panel data model have been obtained and will be presented in the Appendix. But a lot more thinking is needed to establish the asymptotic properties in this case. Also we may consider the derivative estimate of the regression function using the P-splines estimator as well which would induce a wide application in economics.

Note

1. Initially, we consider four kernel-based estimators including the ones proposed by Henderson (2008) and Sun (2009) in the simulation. These two alternatives, however, turn out to be not statistical efficient nor computational efficient. Hence we remove them from our comparison.

Bibliography

[1] Agarwal, Girdhar G., and Studden, W. J., "Asymptotic Integrated Mean Square Error Using Least Squares and Bias Minimizing Splines,", *The Annals of Statistics*, 8 (1980): 1307-1325.

[2] Barrow, D.L. and Smith, P.W., "Asymptotic Properties of Best $L_2[0,1]$ Approximation by Splines with Variable Knots," *Quart. Appl. Math.*, 36 (1978): 293-304.

[3] Chui, C. K., *An Introduction to Wavelets*, Academic Press Professional Inc., 1992.

[4] Claeskens, G., Krivobokova, T. and Opsomer, J. D., "Asymptotic Properties of Penalized Spline Estimators," *Biometrika*, 96 (2009): 529-544.

[5] Doksum, K. and Samarov, A., "Nonparametric Estimation of Global Functionals and A Measure of the Explanatory Power of Covariates in Regression," *The Annuals of Statistics*, 23 (1995): 1443-1473.

[6] Eilers, P.H.C. and Marx, B.D., "Flexible Smoothing Using B-splines and Penalized Likelihood (with Comments and Rejoinder)," *Statistical Science*, 11 (1996): 89-121.

[7] Eilers, P.H.C. and Marx, B.D., "Splines, Knots and Penalties," URL http://www.stat.lsu.edu/faculty/marx/splines_knots_penalties.pdf.

[8] Green, P. J. and Silverman, B. W., *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall, 1994.

[9] Greene, W. B., *Econometric Analysis*, Pearson / Prentice Hall, 7th edition (2012).

[10] Hajargasht, G., "Nonparametric Panel Data Models: A Penalized Spline Approach," Working Paper No. WP05/2009.

[11] Hastie, T. J. and Tibshirani, R. J., *Generalized Additive Models*, Chapman and Hall, London, 1990.

[12] Henderson, D.J., Carroll, R.J., and Li, Q., "Nonparametric Estimation and Testing of Fixed Effects Panel Data Models," *Journal of Econometrics*, 144 (2008): 257-275.

[13] Henderson, D.J. and Ullah, A, "A Nonparametric Random Effects Estimator," *Economics Letters*, 88 (2005): 403-407.

[14] Keele, L., *Semiparametric Regression for the Social Sciences* John Wiley and Sons, 2008.

[15] Krivobokova, T. and Kauermann, G., "A Note on Penalized Spline Smoothing with Correlated Errors," *Journal of the American Statistical Association*, 102 (2007): 1328-1337.

[16] Li, Q. and Racine, J. S., *Nonparametric Econometrics: Theory and Practice*, Princeton and Oxford, 2007.

[17] Li, Y. and Ruppert, D., "On the Asymptotics of Penalized Splines," *Biometrika*, 95 (2008): 415-436.

[18] Lin, X., and Carroll, R. J., "Nonparametric Function Estimation for Clustered Data When the Predictor Is Measured Without/ with Error," *Journal of the American Statistical Association*, 95 (2000): 520-534.

[19] Lin, X., and Carroll, R. J., "Semiparametric Regression for Clustered Data Using Generalized Estimation Equations," *Journal of the American Statistical Association,* 96 (2001): 1045-1056.

[20] Lin, X., and Carroll, R. J., "Semiparametric Estimation in General Repeated Measures Problems," *Journal of the Royal Statistical Society, Series B*, 68 (2006): 68-88.

[21] Lin, X., Wang, N., Welsh, A. H. and Carroll, R. J., "Equvalent Kernels of Smoothing Splines in Nonparametric Regression for Longitudinal/ Clustered Data," *Biometrika*, 91 (2004): 177-194.

[22] Martins-Filho, C. and F. Yao, "Nonparametric Regression Estimation with General Parametric Error Covariance," *Journal of Multivariate Analysis*, 100 (2009): 309-333.

[23] Pagan, A. and Ullah, A. , *Nonparametric Econometrics*, New York: Cambridge, 1999.

[24] Pourahmadi, M., "Joint Mean-Covariance Models with Applications to Longitudinal Data: Unconstraint Parameterisation," *Biometrika*, 86 (1999): 677-690.

[25] Qian, Junhui and Wang, Le, "Estimating Semiparametric Panel Data Models by Marginal Integration," *Journal of Econometrics*, 167 (2012): 483-493.

[26] Ruckstuhl, A. F., Welsh, A. H. and Carroll, R. J., "Nonparametric Function Estimation of the Relationship Between Two Repeatedly Measured Variables," *Statistica Sinica*, 10 (2000): 51-71.

[27] Ruppert, D., Wand, M. P. and Carroll, R. J., *Semiparametric Regression*, Cambridge, 2003.

[28] Schumaker, Larry L., *Spline Functions: Basic Theory*, Cambridge University Press, 3rd edition (2007).

[29] Speckman, P., "Spline Smoothing and Optimal Rates of Convergence in Nonparametric Regression Models," *The Annals of Statistics*, 13 (1985): 970-983.

[30] Stone, C., "Optimal Global Rates of Convergence for Nonparametric Regression," *The Annals of Statistics*, 10 (1982): 1040-1053.

[31] Su, L. and Ullah, A., "Nonparametric and Semiparametric Panel Econometric Models: Estimation and Testing," in *Handbook of Empirical Economics and Finance*, (eds. Ullah, A. and Giles, D. E. A.), Taylor and Francis.

[32] Su, L. and Ullah, A., "Profile Likelihood Estimation of Partially Linear Panel Data Models with Fixed Effects," *Economics Letters*, 92 (2006): 75-81.

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

[34] Su, L., Ullah, A. and Wang, Y., "Nonparametric Regression Estimation with General Parametric Error Covariance: A More Efficient Two-Step Estimator," forthcoming in *Empirical Economics* (2013).

[35] Sun, Yiguo, R. Carroll, and D. Li, "Semiparametric Estimation of Fixed Effects Panel Data Varying Coefficient Models," *Advances in Econometrics*, 25 (2009): 101-129.

[36] Ullah, A and Roy, N., "Nonparametric and Semiparametric Econometrics of Panel Data" in *Handbook of Applied Economics Statistics* (eds. Ullah, A. and Giles, D. E. A.), 1 (1998): 579-604.

[37] Wu, H. and Zhang, J.-T., *Nonparametric Regression Methods for Longitudinal Data Analysis*, Wiley-Interscience, 2006.

[38] Zhou, S., Shen, X. and Wolfe, D. A., "Local Asymptotics for Regression Splines and Confidence Regions," *The Annals of Statistics*, 26 (1998): 1760-1782.

Appendix I-A Technical Support

First recall that the asymptotic properties of generalized least squares (GLS) estimator for linear regression model are straightforward extensions of the properties of ordinary least squares (OLS) by standardizing the model. For our P-splines estimator proposed for random effects panel data model, however, it is somehow more complicated. For example, the asymptotic results of the penalized splines estimator with B-splines basis for cross-sectional data model rely heavily on the band structure of a B-spline design matrix, which is not applicable for random effects panel data model. Therefore a rigorous mathematical proof is necessary to obtain a tenable result for panel data model backdrop. In this Appendix, we provide proofs for Theorem 1-3.

Proof of Theorem 1

For the estimator given in (1.12), we apply the Demmler-Reinsch algorithm (Demmler and Reinsch, 1975) as follows,

Cholesky decomposition:

$\mathbf{N}'\Omega^{-1}\mathbf{N} = \mathbf{R}'\mathbf{R}$, where $\mathbf{R} = (\mathbf{N}'\Omega^{-1}\mathbf{N})^{1/2}$ is an invertible matrix.

Singular value decomposition:

$$\mathbf{R}^{-1\prime}\mathbf{D}\mathbf{R}^{-1} = (\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1/2}\mathbf{D}_q(\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1/2} = \mathbf{U}diag(\mathbf{s})\mathbf{U}',$$

where $\mathbf{U}$ is an orthogonal matrix and $\mathbf{s}$ is the vector of eigenvalues of $(\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1/2}\mathbf{D}_q(\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1/2}$.

Define a matrix $\mathbf{A} \equiv \Omega^{-1/2}\mathbf{N}(\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1/2}\mathbf{U}$. It follows that $\mathbf{A}'\mathbf{A} = \mathbf{I}_{K+p+1}$, and $\mathbf{A}\mathbf{A}' = \Omega^{-1/2}\mathbf{N}(\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1}\mathbf{N}'\Omega^{-1/2}$ is an idempotent matrix. Thus $\mathbf{N}'\Omega^{-1}\mathbf{N} + \lambda\mathbf{D}_q = \mathbf{R}'\mathbf{U}\mathbf{U}^{-1}\mathbf{R} + \lambda\mathbf{R}'\mathbf{U}diag(\mathbf{s})\mathbf{U}^{-1}\mathbf{R} = \mathbf{R}'\mathbf{U}\{\mathbf{I}_{K+p+1} + \lambda diag(\mathbf{s})\}\mathbf{U}^{-1}\mathbf{R}$. The estimator (1.12) can then be written as

$$\begin{aligned}\hat{\mathbf{m}} &= \Omega^{1/2}\Omega^{-1/2}\mathbf{N}\mathbf{R}^{-1}\mathbf{U}\{\mathbf{I}_{K+p+1} + \lambda diag(\mathbf{s})\}^{-1}\mathbf{U}'\mathbf{R}'^{-1}\mathbf{N}'\Omega^{-1/2}\Omega^{-1/2}\mathbf{y} \\ &= \Omega^{1/2}\mathbf{A}\{\mathbf{I}_{K+p+1} + \lambda diag(\mathbf{s})\}^{-1}\mathbf{A}'\Omega^{-1/2}\mathbf{y} \qquad \text{(I-1)}\end{aligned}$$

For later use, we also provide the regression splines estimator as

$$\hat{\mathbf{m}}_{reg} = \mathbf{N}(\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1}\mathbf{N}'\Omega^{-1}\mathbf{y} = \Omega^{1/2}\mathbf{A}\mathbf{A}'\Omega^{-1/2}\mathbf{y} \qquad \text{(I-2)}$$

Hence

$$E(\hat{\mathbf{m}}_{reg}) = \Omega^{1/2}\mathbf{A}\mathbf{A}'\Omega^{-1/2}\mathbf{m}. \qquad \text{(I-3)}$$

The AMSE of $\hat{\mathbf{m}}$ is given as

$$\text{AMSE}(\hat{\mathbf{m}}) = \frac{1}{nT}E\left\{(\hat{\mathbf{m}}^* - \mathbf{m}^*)'\Omega(\hat{\mathbf{m}}^* - \mathbf{m}^*)\right\},$$

where $\hat{\mathbf{m}}^* = \Omega^{-1/2}\hat{\mathbf{m}}$ and $\mathbf{m}^* = \Omega^{-1/2}\mathbf{m}$.

Lemma 1: For any nonnegative matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\lambda^A_{\min} tr(\mathbf{B}) \le tr(\mathbf{A}\mathbf{B}) \le \lambda^A_{\max} tr(\mathbf{B})$$

where $\lambda^A_{\min}$ and $\lambda^A_{\max}$ are respectively the minimum and maximum eigenvalue of matrix $\mathbf{A}$, and $tr(\cdot)$ represents the trace of a matrix.

With Lemma 1 and Ω a matrix with bounded eigenvalues,

$$\text{AMSE}(\hat{\mathbf{m}}) \sim \frac{1}{nT} E\left\{(\hat{\mathbf{m}}^* - \mathbf{m}^*)'(\hat{\mathbf{m}}^* - \mathbf{m}^*)\right\} \tag{I-4}$$

where '$\sim$' means 'being of the same order'.

Then following the scheme proposed by Claeskens (2009), the AMSE in (I-4) can be decomposed as follows:

$$\begin{aligned}\text{AMSE}(\hat{\mathbf{m}}) \sim & \frac{1}{nT}\sum_{k=1}^{K+p+1}\frac{1}{(1+\lambda s_k)^2} + \frac{\lambda^2}{nT}\sum_{k=1}^{K+p+1}\frac{s_k^2 b_k^2}{(1+\lambda s_k)^2} \\ & + \frac{1}{nT}\left[E(\hat{\mathbf{m}}_{reg}) - \mathbf{m}\right]' \Omega^{-1}\left[E(\hat{\mathbf{m}}_{reg}) - \mathbf{m}\right] \qquad \text{(I-5)} \\ & \equiv \text{AMSE}_1 + \text{AMSE}_2 + \text{AMSE}_3\end{aligned}$$

where $b_1, \cdots, b_{K+p+1}$ are elements of $\mathbf{b} = \mathbf{A}'\mathbf{m}^* = \mathbf{U}'(\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1/2}\mathbf{N}'\Omega^{-1}\mathbf{m}$.

Lemma 2 (Speckman, 1985):

Under (A2) and for the eigenvalues defined in (I-1),

$$s_1 = s_2 = \cdots = s_q = 0,$$

$$s_k = n^{-1}(k-q)^{2q} c\{1 + o_p(1)\},\ k = q+1, \cdots, K+p+1$$

where c is some constant that depends only on q and ρ.

For notational simplicity, define $K_q = (K+p+1-q)(\lambda c\{1+o_p(1)\})^{1/2q} n^{-1/2q}$ such that $s_{K+p+1} = K_q^{2q}$. It has been verified that the asymptotic scenario would be

similar to that of regression splines when $K_q < 1$, and similar to that of smoothing splines when $K_q > 1$. It can also be shown that the AMSE achieves the optimal rate for $K_q < 1$, but not for $K_q > 1$. Therefore we focus only on $K_q < 1$ case for the proof.

Following the method presented in Claeskens (2009), we obtain

$$\text{AMSE}_1 \sim O_p\left(\frac{K}{n}\right). \tag{I-6}$$

To obtain the order of AMSE_2, we need the following lemmas.

Lemma 3 (Cardot, 2000): $\| \mathbf{D}_q \|_\infty = O_p(\delta^{-2q+1})$.

Lemma 4: $\| \mathbf{G}_{K,n} \|_\infty = O_p(\delta)$, where $\mathbf{G}_{K,n} \equiv \dfrac{\mathbf{N}'\Omega^{-1}\mathbf{N}}{nT}$.

Lemma 5: $\| \mathbf{G}_{K,n}^{-1} \|_\infty = O_p(\delta^{-1})$.

Proof of Lemma 5: To prove Lemma 5, we need Lemma 6.

Lemma 6: Let $\mathbf{V}$ be a positive definite $n \times n$ matrix, and let $\mathbf{X}$ be an $n \times k$ matrix with full rank of k. Then

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \geq (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}.$$

From Lemma 6, we have

$$\left(\frac{\mathbf{N}'\Omega^{-1}\mathbf{N}}{nT}\right)^{-1} \leq \left(\frac{\mathbf{N}'\mathbf{N}}{nT}\right)^{-1}\left(\frac{\mathbf{N}'\Omega\mathbf{N}}{nT}\right)\left(\frac{\mathbf{N}'\mathbf{N}}{nT}\right)^{-1}.$$

Using Lemma 6.3 stated in Zhou (1998) that $\| (\frac{\mathbf{N}'\mathbf{N}}{nT})^{-1} \|_\infty = O_p(\delta^{-1})$, along with the fact that $\| \frac{\mathbf{N}'\Omega\mathbf{N}}{nT} \|_\infty = O_p(\delta)$, we prove Lemma 5.

From the singular value decomposition stated above, we have

$$\sum_{k=1}^{K+p+1} b_k^2 s_k^2 = \beta_m{}' \mathbf{D}_q (\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1} \mathbf{D}_q \beta_m, \tag{I-7}$$

where $\beta_m = (\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1}\mathbf{N}'\Omega^{-1}\mathbf{m}$. Then the following inequality holds with Lemma 1,

$$\frac{1}{(1+K_q^{2q})^2}\beta_m{}' \mathbf{D}_q (\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1} \mathbf{D}_q \beta_m \leq \sum_{k=1}^{K+p+1} \frac{b_k^2 s_k^2}{(1+\lambda s_k)^2} \leq \beta_m{}' \mathbf{D}_q (\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1} \mathbf{D}_q \beta_m .$$

Hence

$$\sum_{k=1}^{K+p+1} \frac{b_k^2 s_k^2}{(1+\lambda s_k)^2} \sim \beta_m{}' \mathbf{D}_q (\mathbf{N}'\Omega^{-1}\mathbf{N})^{-1} \mathbf{D}_q \beta_m .$$

Under (R3) stated in Claeskens (2009) and with Lemma 3 and Lemma 5, we obtain

$$\sum_{k=1}^{K+p+1} \frac{b_k^2 s_k^2}{(1+\lambda s_k)^2} \sim O_p(\frac{K^{2q}}{n}),$$

and therefore,

$$\text{AMSE}_2 \sim O_p(\frac{\lambda^2 K^{2q}}{n^2}). \tag{I-8}$$

To obtain the order of AMSE_3, we need lemma 6.10 in Agarwal and Studden (1980). It follows that $E[\hat{m}_{reg}(x)] - m(x) = O_p(\delta^{p+1})$, from which we get

$$\text{AMSE}_3 = O_p(K^{-2(p+1)}). \tag{I-9}$$

From (I-6), (I-8) and (I-9), the order of AMSE is obtained as

$$\text{AMSE} = O_p(\frac{K}{n}) + O_p(\frac{\lambda^2 K^{2q}}{n^2}) + O_p(K^{-2(p+1)}). \tag{I-10}$$

By matching the first term and third term, we have $K = O_p(n^{1/(2p+3)})$. After matching the second term, we get $\lambda = O_p(n^j)$, where $j \le \frac{p+2-q}{2p+3}$. Upon this the optimal rate of convergence is achieved as $n^{-(2p+2)/(2p+3)}$ (see Stone, 1982).

Proof of Theorem 2

For the B-splines matrix $\mathbf{N}_{nT\times(K+p+1)}$, we denote it as $\mathbf{N} = (\mathbf{N}_1', \cdots, \mathbf{N}_n')'$, where $\mathbf{N}_i = \begin{bmatrix} N_{-p}(x_{i1}) & \cdots & N_K(x_{i1}) \\ \vdots & \ddots & \vdots \\ N_{-p}(x_{iT}) & \cdots & N_K(x_{iT}) \end{bmatrix}$. And $\Omega^{-1} = \mathbf{I}_n \otimes \Sigma^{-1}$, where $\Sigma^{-1} = (\sigma^{ts})_{t,s=1}^{T}$.

Lemma 7: $\| \mathbf{G}_{K,n} - \mathbf{G} \|_\infty = o_p(\delta)$, where $\mathbf{G}$ is an $nT \times nT$ matrix with (j_0, l_0)th element being

$$(\mathbf{G})_{j_0,l_0} = \frac{1}{T}\sum_{t=1}^{T}\int_a^b \sigma^{tt} N_j(x_{it}) N_l(x_{it}) dQ^{[t]}(\cdot)$$

$$+\frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{s=1 \\ s\neq t}}^{T}\int_a^b\int_a^b \sigma^{ts} N_j(x_{it}) N_l(x_{is}) dQ^{[s]}(\cdot) dQ^{[t]}(\cdot),$$

where $j_0 = j + p + 1$, and $l_0 = l + p + 1$.

Proof of Lemma 7: With the notations stated above, the elements of $\mathbf{G}_{K,n}$ are

$$(\mathbf{G}_{K,n})_{j_0,l_0} = \frac{1}{T}\sum_{t=1}^{T}\int_a^b \sigma^{tt} N_j(x_{it}) N_l(x_{it}) dQ_n^{[t]}(\cdot)$$

$$+\frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{s=1 \\ s\neq t}}^{T}\int_a^b\int_a^b \sigma^{ts} N_j(x_{it}) N_l(x_{is}) dQ_n^{[s]}(\cdot) dQ_n^{[t]}(\cdot).$$

Hence,

$$|(\mathbf{G}_{K,n})_{j_0,l_0} - (\mathbf{G})_{j_0,l_0}| \le |\frac{1}{T}\sum_{t=1}^{T}\int_a^b \sigma^{tt} N_j(x_{it}) N_l(x_{it}) d(Q_n^{[t]} - Q^{[t]})(\cdot)|$$

$$+ |\frac{1}{T}\sum_{t=1}^{T}[\int_a^b N_j(x_{it}) \sum_{\substack{s=1 \\ s\neq t}}^{T} \sigma^{ts} \int_a^b N_l(x_{is}) dQ^{[s]}(\cdot) d(Q_n^{[t]} - Q^{[t]})(\cdot) + \int_a^b N_j(x_{it}) o_p(\delta) dQ_n^{[t]}(\cdot)]|$$

$$= o_p(\delta) + o_p(\delta^2) = o_p(\delta).$$

Lemma 8: $\| \mathbf{G}_{K,n}^{-1} - \mathbf{G}^{-1} \|_{\infty} = O_p(\delta^{-1})$.

The P-splines estimator can be written as follows:

$$\hat{m}(x) = N'(x)(\mathbf{G}_{K,n} + \frac{\lambda}{nT}\mathbf{D}_q)^{-1} \frac{\mathbf{N}'}{nT}\Omega^{-1}\mathbf{y}$$

$$= N'(x)(\mathbf{I}_{K+p+1} + \frac{\lambda}{nT}\mathbf{G}_{K,n}^{-1}\mathbf{D}_q)^{-1}\mathbf{G}_{K,n}^{-1} \frac{\mathbf{N}'}{nT}\Omega^{-1}\mathbf{y}.$$

Since $K_q = \lambda_{\max}^{(\frac{\lambda}{nT}\mathbf{G}_{K,n}^{-1}\mathbf{D}_q)} < 1$, we have

$$(\mathbf{I}_{K+p+1} + \frac{\lambda}{nT}\mathbf{G}_{K,n}^{-1}\mathbf{D}_q)^{-1} = \mathbf{I}_{K+p+1} - \frac{\lambda}{nT}\mathbf{G}_{K,n}^{-1}\mathbf{D}_q + (\frac{\lambda}{nT}\mathbf{G}_{K,n}^{-1}\mathbf{D}_q)^2 - \cdots$$

Hence

$$\hat{m}(x) = \hat{m}_{reg}(x) - \frac{\lambda}{nT}N'(x)\mathbf{G}_{K,n}^{-1}\mathbf{D}_q\mathbf{G}_{K,n}^{-1}\frac{\mathbf{N}'}{nT}\Omega^{-1}\mathbf{y} + r_n.$$

It follows that the asymptotic bias is

$$E[\hat{m}(x) \mid \mathbf{x}] - m(x) = E[\hat{m}_{reg}(x) \mid \mathbf{x}] - m(x) - \frac{\lambda}{nT}N'(x)\mathbf{G}_{K,n}^{-1}\mathbf{D}_q\mathbf{G}_{K,n}^{-1}\frac{\mathbf{N}'}{nT}\Omega^{-1}\mathbf{m} + E[r_n],$$

where $E[r_n]$ satisfies

$$| E[r_n] | \le \frac{\lambda K_q^{2q}}{nT(1+K_q^{2q})} | N'(x)\mathbf{G}_{K,n}^{-1}\mathbf{D}_q\mathbf{G}_{K,n}^{-1}\mathbf{N}'\frac{1}{nT}\Omega^{-1}\mathbf{m} |.$$

Following Zhou (1998) and Lemma 3 in Claeskens (2009), we have

$$E[\hat{m}(x) \mid \mathbf{x}] - m(x) = O_p(\delta^{p+1}) + O_p(\lambda n^{-1}\delta^{-q}).$$

With $\delta = O(K^{-1})$ and the optimal rate of K and λ given above, it turns out that

$$E[\hat{m}(x) \mid \mathbf{x}] - m(x) = O_p(n^{-(p+1)/(2p+3)}),$$

achieves the optimal rate of convergence.

Similarly, we can obtain the asymptotic variance

$$Var(\hat{m}(x) \mid \mathbf{x}) = Var(\hat{m}_{reg}(x) \mid \mathbf{x}) - \frac{\lambda c}{(nT)^2} N'(x)\mathbf{G}_{K,n}^{-1}\mathbf{D}_q\mathbf{G}_{K,n}^{-1}N(x)$$

$$-\frac{\lambda c}{(nT)^2} N'(x)\mathbf{G}^{-1}\mathbf{D}_q\mathbf{G}^{-1}N(x) - \frac{\lambda c}{(nT)^2} N'(x)(\mathbf{G}_{K,n}^{-1}\mathbf{D}_q\mathbf{G}_{K,n}^{-1} - \mathbf{G}^{-1}\mathbf{D}_q\mathbf{G}^{-1})N(x).$$

Following Zhou (1998) and Lemma 2 in Claeskens (2009), we have

$$Var(\hat{m}(x) \mid \mathbf{x}) = \frac{1}{nT} N'(x)\mathbf{G}^{-1}N(x) - \frac{\lambda c}{(nT)^2} N'(x)\mathbf{G}^{-1}\mathbf{D}_q\mathbf{G}^{-1}N(x)$$

$$+o_p(n^{-1}\delta^{-1}) + o_p(\lambda n^{-2}\delta^{-(2q+1)}),$$

and it follows that $Var(\hat{m}(x) \mid \mathbf{x}) = O_p(n^{-1}\delta^{-1}) + O_p(\lambda n^{-2}\delta^{-(2q+1)}) = O_p(n^{-2(p+1)/(2p+3)})$.

Proof of Theorem 3

Since $\frac{E[\hat{m}(x)\,|\,\mathbf{x}]-[m(x)+b(x)]}{\sqrt{Var(\hat{m}(x)\,|\,\mathbf{x})}}=o_p(1)$, it suffices to prove

$\frac{\hat{m}(x)-E[\hat{m}(x)\,|\,\mathbf{x}]}{\sqrt{Var(\hat{m}(x)\,|\,\mathbf{x})}}\xrightarrow{d}N(0,1)$, which holds following Liapunov Central Limit Theorem under the assumption that $K=O(n^{1/2})$.

Appendix I-B Simulation Results for FEPDM

AMSE of function estimate for fixed effects panel data model (c=1)

DGP	n	Su (2006)	Henderson (2008)	Qian	P-Splines
1	50	.0787	.1094	.0724	.0591
	100	.0552	.0412	.0386	.0163
	200	.0352	.0199	.0229	.0162
2	50	.1121	.1413	.2834	.0698
	100	.0621	.0889	.1352	.0484
	200	.0339	.0446	.0701	.0250
3	50	.0897	.3603	.3745	.0857
	100	.0547	.2347	.2869	.0476
	200	.0352	.1982	.2532	.0249

Chapter 3

Asymptotic Variance-Covariance Matrices of Two-Stage Estimators for Simultaneous Equations System

3.1 Introduction

On occasions we may come across joint determination of two or more variables by a system of equations. Such multiple equations framework is often referred to as simultaneous equations system or interdependent equations model. A simple classical example is the partial equilibrium model of supply and demand of a market good where the equilibrium price and quantity are simultaneously determined. Possible solutions for simultaneous equations system include two-stage estimation, generalized method of moments (GMM) and full information maximum likelihood (FIML) among others. FIML, as a joint method, is famous for its computational complexity and high implemental cost (Blanchard, 1983). It may be unattractive in

some other sense as well, see Murphy and Topel (1985). For GMM method, we will discuss it shortly at the end of this section. Our main interest of this work, on the other hand, is the two-stage estimation.

Due to its computational simplicity and instrumental variable interpretation, two-stage method to estimate simultaneous equations model has been quite popular among applied econometricians during the last few decades (Greene, 2012). It is the subject of a long literature, such as two-stage least squares (Brundy and Jorgenson, 1971), two-step estimator for the correction of selectivity bias (Lee, 1976; Heckman, 1979; Tunali, 1986), two-stage probit (Lee, Maddala and Trost, 1980) and two-stage bivariate probit (Mohanty, 1995; Heywood and Mohanty, 2005). The common approach of all these two-stage procedures is to first estimate the reduced form equations of endogenous variables by appropriate econometric methods, and then use their predicted values as instruments in the second stage estimation of the structural equations. This avoids complex estimation and yields consistent coefficient estimates under fairly general conditions, and therefore is widely used by numerous researchers in applied economics.

Despite the consistency for coefficient estimates, it is well known that many practitioners employ the two-step procedure without correcting the second-step estimated standard errors and may therefore generate unreliable test statistics. In other words, two-stage estimation for statistical inference purposes requires derivation of

corrected asymptotic variance-covariance matrices of the second stage estimator. These derivations that often times involve sophisticated statistical and mathematical exercises may be tedious and may therefore be quite daunting to practitioners of econometrics with little expertise in that area. Fortunately, formulas for the covariance matrices of the second-stage estimators in some special cases have been derived by their proponents and some of them have been built into standard statistical software packages available today. Variance-covariance matrices of general two-stage estimators, however, are not yet available in the current literature. For the benefit of numerous applied economists, this chapter develops the asymptotic variance-covariance matrices of two-stage estimators when the simultaneous equations model consists of a mixture of continuous and binary dependent variables.

Simultaneous equations models that involve both continuous and binary endogenous variables are quite common in applied econometric studies. See, for example, the recent work by Lydon and Chevalier (2002), Mohanty (2009a; 2009b) and Mohanty and Ullah (2012) for varieties of these models. Following Amemiya (1978), Maddala (1983) outlined two-stage procedures to estimate several simultaneous equations models with two dependent variables, at least one of them binary, and derived their asymptotic variance-covariance matrices. Here we consider a general simultaneous equations model with any number of dependent variables that include a mixture of binary and continuous endogenous variables, and derive general formulas

for asymptotic covariance matrices of resulting two stage estimators. We use a specific example to show how practitioners can easily write asymptotic variance-covariance matrices of their two-stage estimators using our general formulas. Although no novelty is claimed with regard to the derivation of our results, we strongly believe that formulas derived in this study would be of considerable help to practitioners of econometrics who may need them for their applied work, but may not have the ability to derive them for themselves (Mohanty, 2009b). To that extent, this study contributes to the literature as a useful practitioner's tool.

It is also important to note that, for a simultaneous equations system, an alternative to two-stage estimation would be the GMM estimation (Wooldridge, 2010). However, a close scrutiny of the model proposed in this study reveals a different layer of complexity in applying GMM to this model. To see this, we present in this paragraph two possible model specifications based on how the binary endogenous variables enter the system. First, if all the binary endogenous variables enter the system as latent continuous variables only, the parameters of the model cannot be estimated by GMM without following a multi-step procedure as the latent variables are unobserved (Windmeijer and Santos Silva, 1997). Alternatively, if the binary variables themselves appear as explanatory variables in the model, the orthogonal conditions required by GMM may be constructed for recursive models only (Bierens,

2007). For general non-recursive models proposed in this study, however, a GMM estimator, to the knowledge of the authors, has not yet been proposed in the literature.

The rest of this chapter is organized as follows. Section 3.2 presents the general simultaneous equations model and its two stage estimators. In Section 3.3, we provide the general formula for the asymptotic variance-covariance matrices of these two-stage estimators. Using data from the National Longitudinal Survey of Youth (1979), Section 3.4 presents an empirical application which shows how our procedure can be used to write the covariance matrices in specific models. The final section 3.5 concludes.

3.2 Two-Stage Estimators for Simultaneous Equations System

A two-stage estimator is provided in this section for a simultaneous equation system with M continuous and R binary dependent variables, where M and R are arbitrary nonnegative integers.

Define $M+R$ observed endogenous variables $y_1, y_2, \cdots, y_M, y_{M+1}, \cdots, y_{M+R}$, where $y_1, \cdots, y_M$ are continuous and $y_{M+1}, \cdots, y_{M+R}$ are binary. Each binary variable y_{M+r} $(r = 1, \cdots, R)$ is observed only when the latent continuous variable y_{M+r}^* is positive. The general structural simultaneous equations model can be written as

$$y_m = Y_{(-m)}\gamma_m + Y^*\delta_m + X_m\beta_m + \varepsilon_m, \qquad m = 1,\cdots,M \tag{3.1}$$

$$y^*_{M+r} = Y\gamma_{M+r} + Y^*_{(-r)}\delta_{M+r} + X_{M+r}\beta_{M+r} + \varepsilon_{M+r}, \qquad r = 1,\cdots,R \tag{3.2}$$

$$y_{M+r} = 1, \quad if \ \ y^*_{M+r} > 0; \ \ = 0, \ \ otherwise, \tag{3.3}$$

where each of $y_1,\cdots,y_{M+R},y^*_{M+1},\cdots,y^*_{M+R}$ is an $n\times 1$ vector, $Y=(y_1,y_2,\cdots,y_M)$, $Y_{(-m)}=(y_1,\cdots,y_{m-1},y_{m+1},\cdots,y_M)$, $Y^*=(y^*_{M+1},y^*_{M+2},\cdots,y^*_{M+R})$, $Y^*_{(-r)}=(y^*_{M+1},\cdots,y^*_{M+r-1},y^*_{M+r+1},\cdots,y^*_{M+R})$, $\gamma_m=(\gamma_{m1},\cdots,\gamma_{m,m-1},\gamma_{m,m+1},\cdots,\gamma_{mM})'$, $\delta_m=(\delta_{m1},\delta_{m2},\cdots,\delta_{mR})'$, $\gamma_{M+r}=(\gamma_{M+r,1},\gamma_{M+r,2},\cdots,\gamma_{M+r,M})'$, $\delta_{M+r}=(\delta_{M+r,1},\cdots,\delta_{M+r,r-1},\delta_{M+r,r+1},\cdots,\delta_{M+r,R})'$, and $X_1,\cdots,X_{M+R}$ are $n\times K_1,\cdots,n\times K_{M+R}$ matrices of exogenous variables respectively, and the error terms $\varepsilon's$ follow the standard assumptions of classical regression models.

The reduced form equations of (3.1) – (3.2) are

$$y_m = X\pi_m + v_m, \qquad m = 1,\cdots,M \tag{3.4}$$

$$y^*_{M+r} = X\pi_{M+r} + v_{M+r}, \qquad r = 1,\cdots,R \tag{3.5}$$

where X is an $n\times K$ matrix of all distinct exogenous variables obtained from $X_1,\cdots,X_{M+R}$, and $\pi_1,\cdots,\pi_{M+R}$ are the reduced form parameters. Note that $v_1=(v_{11},\cdots,v_{1n})',\cdots,v_{M+R}=(v_{M+R,1},\cdots,v_{M+R,n})'$ are error terms whose variance-covariance structures are defined as follows[1]:

$$
\begin{aligned}
&V(v_{mi}) = \sigma_m^2, && m = 1, \cdots, M \\
&V(v_{M+r,i}) = 1, && r = 1, \cdots, R \\
&Cov(v_{si}, v_{ti}) = \sigma_{st}, && s, t = 1, \cdots, M+R \ \ and \ \ s \neq t \\
&Cov(v_{si}, v_{tj}) = 0, && i \neq j
\end{aligned}
\qquad (3.6)
$$

In the first stage, the reduced form parameters $\pi_m's$ $(m = 1, \cdots, M)$ are estimated by OLS,

$$\hat{\pi}_m = (X'X)^{-1} X'y_m, \qquad (3.7)$$

and $\tilde{\pi}_{M+r}'s$ $(r = 1, \cdots, R)$, the maximum likelihood estimators of $\pi_{M+r}'s$, are obtained by maximizing the following likelihood function:

$$L(\pi_{M+r}) = \prod_{i=1}^{n} \left\{ F(x_i'\pi_{M+r}) \right\}^{y_{M+r,i}} \left\{ 1 - F(x_i'\pi_{M+r}) \right\}^{1-y_{M+r,i}}, \qquad (3.8)$$

where x_i' is the i^{th} row of matrix X, and $F(x_i'\pi_{M+r})$ is the cumulative distribution function of v_{M+r}. Throughout this paper, we assume $v_{M+r}'s$ follow normal distribution with the variance-covariance structure defined above.

In the second stage of estimation, the first structural equation following Maddala (1983) and Amemiya (1979) can be written as

$$
\begin{aligned}
y_1 &= [X\pi_2 + v_2, \cdots, X\pi_M + v_M, X\pi_{M+1} + v_{M+1}, \cdots, X\pi_{M+R} + v_{M+R}][\gamma_1' \ \ \delta_1']' + X_1\beta_1 + \varepsilon_1 \\
&= X[\hat{\pi}_2, \cdots, \hat{\pi}_M, \tilde{\pi}_{M+1}, \cdots, \tilde{\pi}_{M+R}, J_1][\gamma_1' \delta_1' \beta_1']' + w_1 \\
&= X[\hat{\pi}_{(-1)} \ \tilde{\pi} \ J_1][\gamma_1' \delta_1' \beta_1']' + w_1 \\
&= X\hat{H}_1 b_1 + w_1,
\end{aligned}
$$

where $w_1 = v_1 - X[\hat{\pi}_2 - \pi_2, \cdots, \hat{\pi}_M - \pi_M, \tilde{\pi}_{M+1} - \pi_{M+1}, \cdots, \tilde{\pi}_{M+R} - \pi_{M+R}][\gamma_1', \delta_1']'$,

$XJ_1 = X_1$, and $J_1 = [I_{K_1} \; 0_{K_1 \times (K-K_1)}]'$. A more general formulation for the first set of *m* equations similar to the above y_1 equation is as follows:

$$y_m = X\hat{H}_m b_m + w_m, \tag{3.9}$$

where $b_m = (\gamma_m' \; \delta_m' \; \beta_m')'$; $\hat{H}_m = [\hat{\pi}_{(-m)} \; \tilde{\pi} \; J_m]$,

in which $\hat{\pi}_{(-m)} = (\hat{\pi}_1, \cdots, \hat{\pi}_{m-1}, \hat{\pi}_{m+1}, \cdots, \hat{\pi}_M)$, $\tilde{\pi} = (\tilde{\pi}_{M+1}, \cdots, \tilde{\pi}_{M+R})$, and

$J_m = [I_{K_M} 0_{K \times (K-K_M)}]'$ is a matrix that consists of 1's and 0's to select variables in X_m from X. Also, $w_m = v_m - X[\hat{\pi}_{(-m)} - \pi_{(-m)}^{(1)} \; \tilde{\pi} - \pi^{(2)}][\gamma_m' \; \delta_m']'$, where

$\pi_{(-m)}^{(1)} = (\pi_1, \cdots, \pi_{m-1}, \pi_{m+1}, \cdots, \pi_M)$ and $\pi^{(2)} = (\pi_{M+1}, \cdots, \pi_{M+R})$. The OLS estimator of b_m therefore is written as $\hat{b}_m = (\hat{H}_m' X' X \hat{H}_m)^{-1} \hat{H}_m' X' y_m, \quad m = 1, \cdots, M.$

Next, we estimate $b_{M+r} = (\gamma_{M+r}' \; \delta_{M+r}' \; \beta_{M+r}')', \quad r = 1, \cdots, R$. Following the previous scheme and notations, we have

$$\begin{aligned} y_{M+r}^* &= [X\pi_1 + v_1, \cdots, X\pi_{M+r-1} + v_{M+r-1}, X\pi_{M+r+1} + v_{M+r+1}, \cdots, X\pi_{M+R} + v_{M+R}] \begin{bmatrix} \gamma_{M+r} \\ \delta_{M+r} \end{bmatrix} \\ &\quad + X_{M+r}\beta_{M+r} + \varepsilon_{M+r} \\ &= X[\pi_{(-(M+r))} \; J_{M+r}][\gamma_{M+r}' \; \delta_{M+r}' \; \beta_{M+r}']' + v_{M+r} \\ &\equiv XG^{(M+r)} b_{M+r} + v_{M+r}, \end{aligned} \tag{3.10}$$

Comparing equation (3.10) with equation (3.5), we have

$$\pi_{M+r} = G^{(M+r)} b_{M+r}. \tag{3.11}$$

Thus the likelihood function can be written as

$$L(\pi_{M+r}) = L(\pi_{(-(M+r))}, b_{M+r})$$

$$= \prod_{i=1}^{n} \left\{ F(x_i' G^{(M+r)} b_{M+r}) \right\}^{y_{M+r,i}} \left\{ 1 - F(x_i' G^{(M+r)} b_{M+r}) \right\}^{1-y_{M+r,i}}. \qquad (3.12)$$

The MLE $\tilde{b}_{M+r}$ of b_{M+r} is then obtained by maximizing $L(\hat{\pi}, \tilde{\pi}_{(-(M+r))}, b_{M+r})$.

3.3 Asymptotic Variance-Covariance Matrices for the Two-Stage Estimators

In this section, we derive the general formulas for asymptotic variance-covariance matrices of the two-stage estimators $\hat{b}_m$ and $\tilde{b}_{M+r}$. Note that

$$\hat{b}_m - b_m = (\hat{H}'_m X' X \hat{H}_m)^{-1} \hat{H}'_m X' w_m,$$

and consequently,

$$V(\hat{b}_m) = (\hat{H}'_m X' X \hat{H}_m)^{-1} \hat{H}'_m X' V(w_m) X \hat{H}_m (\hat{H}'_m X' X \hat{H}_m)^{-1}. \qquad (3.13)$$

To get $V(\hat{b}_m)$ therefore we need $V(w_m)$ which requires the asymptotic distribution of $\hat{\pi}_{(-m)}$ and $\tilde{\pi}$. The former is easy to obtain from (3.6) and (3.7). For $\tilde{\pi}_{M+r}$, we note from (3.8) that

$$\tilde{\pi}_{M+r} - \pi_{M+r} \cong I^{-1}(\pi_{M+r}) S(\pi_{M+r}), \qquad (3.14)$$

where $I(\pi_{M+r}) = -E\left(\frac{\partial^2 \log L}{\partial \pi_{M+r} \partial \pi'_{M+r}}\right)$, $S(\pi_{M+r}) = \frac{\partial \log L}{\partial \pi_{M+r}}$, and both sides of (3.14) have the same asymptotic distribution. Also $V(\tilde{\pi}_{M+r}) = I^{-1}(\pi_{M+r})$. From equation (3.9), we have

$$V(w_m) = V(v_m - X[\hat{\pi}_{(-m)} - \pi^{(1)}_{(-m)} \;\; \tilde{\pi} - \pi^{(2)}] \begin{bmatrix} \gamma_m \\ \delta_m \end{bmatrix})$$

$$= V(v_m - \gamma_{m1} X(\hat{\pi}_1 - \pi_1) - \cdots - \gamma_{m,m-1} X(\hat{\pi}_{m-1} - \pi_{m-1}) - \gamma_{m,m+1} X(\hat{\pi}_{m+1} - \pi_{m+1})$$
$$- \cdots - \gamma_{mM} X(\hat{\pi}_M - \pi_M) - \delta_{m1} X(\tilde{\pi}_{M+1} - \pi_{M+1}) - \cdots - \delta_{mR} X(\tilde{\pi}_{M+R} - \pi_{M+R}))$$

$$\begin{aligned}
&= V(v_m) \\
&+ \sum_{\substack{s=1 \\ s \neq m}}^{M} V(\gamma_{ms} X(\hat{\pi}_s - \pi_s)) \\
&+ \sum_{r=1}^{R} V(\delta_{mr} X(\tilde{\pi}_{M+r} - \pi_{M+r})) \\
&- \sum_{\substack{s=1 \\ s \neq m}}^{M} \left[Cov(v_m, \gamma_{ms} X(\hat{\pi}_s - \pi_s)) + Cov(\gamma_{ms} X(\hat{\pi}_s - \pi_s), v_m) \right] \\
&- \sum_{r=1}^{R} \left[Cov(v_m, \delta_{mr} X(\tilde{\pi}_{M+r} - \pi_{M+r})) + Cov(\delta_{mr} X(\tilde{\pi}_{M+r} - \pi_{M+r}), v_m) \right] \\
&+ \frac{1}{2} \sum_{\substack{s=1 \\ s \neq m}}^{M} \sum_{\substack{t=1 \\ t \neq m \\ t \neq s}}^{M} \begin{bmatrix} Cov(\gamma_{ms} X(\hat{\pi}_s - \pi_s), \gamma_{mt} X(\hat{\pi}_t - \pi_t)) \\ + Cov(\gamma_{mt} X(\hat{\pi}_t - \pi_t), \gamma_{ms} X(\hat{\pi}_s - \pi_s)) \end{bmatrix} \\
&+ \frac{1}{2} \sum_{r_1=1}^{R} \sum_{\substack{r_2=1 \\ r_2 \neq r_1}}^{R} \begin{bmatrix} Cov(\delta_{mr_1} X(\tilde{\pi}_{M+r_1} - \pi_{M+r_1}), \delta_{mr_2} X(\tilde{\pi}_{M+r_2} - \pi_{M+r_2})) \\ + Cov(\delta_{mr_2} X(\tilde{\pi}_{M+r_2} - \pi_{M+r_2}), \delta_{mr_1} X(\tilde{\pi}_{M+r_1} - \pi_{M+r_1})) \end{bmatrix} \\
&+ \sum_{\substack{s=1 \\ s \neq m}}^{M} \sum_{r=1}^{R} \begin{bmatrix} Cov(\gamma_{ms} X(\hat{\pi}_s - \pi_s), \delta_{mr} X(\tilde{\pi}_{M+r} - \pi_{M+r})) \\ + Cov(\delta_{mr} X(\tilde{\pi}_{M+r} - \pi_{M+r}), \gamma_{ms} X(\hat{\pi}_s - \pi_s)) \end{bmatrix}
\end{aligned} \tag{3.15}$$

By splitting $V(w_m)$ into eight parts, we can calculate each separately. For notational convenience, we call them (*1) through (*8), and they are reported in Appendix II-A.

After further simplifications shown in Appendix II-A, we derive from equation (3.13) the following asymptotic covariance matrix of $\hat{b}_m$:

$$V(\hat{b}_m) = c_m(\hat{H}_m'X'X\hat{H}_m)^{-1} + (\hat{H}_m'X'X\hat{H}_m)^{-1}\hat{H}_m'X'XV_{0m}X'X\hat{H}_m(\hat{H}_m'X'X\hat{H}_m)^{-1}. \qquad (3.16)$$

The matrix V_{0m} and the scalar c_m are defined in the Appendix.

Hereafter, we devote to the derivation of $V(\tilde{b}_{M+r})$. Similar to equation (3.14), we have

$$\tilde{b}_{M+r} - b_{M+r} = (E\frac{-\partial^2 \log L}{\partial b_{M+r}\partial b_{M+r}'})^{-1}\begin{bmatrix} \frac{\partial \log L}{\partial b_{M+r}} + \sum_{m=1}^{M} E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi_m'}\right)(\hat{\pi}_m - \pi_m) \\ +\sum_{\substack{t=1 \\ t\neq r}}^{R} E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi_{M+t}'}\right)(\tilde{\pi}_{M+t} - \pi_{M+t}) \end{bmatrix}$$

$$\equiv (E\frac{-\partial^2 \log L}{\partial b_{M+r}\partial b_{M+r}'})^{-1}\Lambda_{M+r}. \qquad (3.17)$$

Thus

$$V(\tilde{b}_{M+r}) = (E\frac{\partial^2 \log L}{\partial b_{M+r}\partial b'_{M+r}})^{-1} V(\Lambda_{M+r})(E\frac{\partial^2 \log L}{\partial b_{M+r}\partial b'_{M+r}})^{-1}, \tag{3.18}$$

where

$$\begin{aligned}
V(\Lambda_{M+r}) &= V\left(\frac{\partial \log L}{\partial b_{M+r}}\right) \\
&+ \sum_{m=1}^{M} E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_m}\right) V(\hat{\pi}_m - \pi_m) E\left(\frac{\partial^2 \log L}{\partial \pi_m \partial b'_{M+r}}\right) \\
&+ \sum_{\substack{t=1 \\ t\neq r}}^{R} E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{M+t}}\right) V(\tilde{\pi}_{M+t} - \pi_{M+t}) E\left(\frac{\partial^2 \log L}{\partial \pi_{M+t} \partial b'_{M+r}}\right) \\
&+ \sum_{m=1}^{M} \begin{bmatrix} Cov(\frac{\partial \log L}{\partial b_{M+r}}, E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_m}\right)(\hat{\pi}_m - \pi_m)) \\ +Cov(E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_m}\right)(\hat{\pi}_m - \pi_m), \frac{\partial \log L}{\partial b_{M+r}}) \end{bmatrix} \\
&+ \sum_{\substack{t=1 \\ t\neq r}}^{R} \begin{bmatrix} Cov(\frac{\partial \log L}{\partial b_{M+r}}, E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{M+t}}\right)(\tilde{\pi}_{M+t} - \pi_{M+t})) \\ +Cov(E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{M+t}}\right)(\hat{\pi}_{M+t} - \pi_{M+t}), \frac{\partial \log L}{\partial b_{M+r}}) \end{bmatrix} \\
&+ \frac{1}{2}\sum_{m_1=1}^{M}\sum_{\substack{m_2=1 \\ m_2\neq m_1}}^{M} \begin{bmatrix} Cov(E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{m_1}}\right)(\hat{\pi}_{m_1} - \pi_{m_1}), E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{m_2}}\right)(\hat{\pi}_{m_2} - \pi_{m_2})) \\ +Cov(E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{m_2}}\right)(\hat{\pi}_{m_2} - \pi_{m_2}), E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{m_1}}\right)(\hat{\pi}_{m_1} - \pi_{m_1})) \end{bmatrix} \\
&+ \frac{1}{2}\sum_{\substack{s=1 \\ s\neq r}}^{R}\sum_{\substack{t=1 \\ t\neq r \\ t\neq s}}^{R} \begin{bmatrix} Cov(E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{M+s}}\right)(\tilde{\pi}_{M+s} - \pi_{M+s}), E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{M+t}}\right)(\tilde{\pi}_{M+t} - \pi_{M+t})) \\ +Cov(E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{M+t}}\right)(\tilde{\pi}_{M+t} - \pi_{M+t}), E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{M+s}}\right)(\tilde{\pi}_{M+s} - \pi_{M+s})) \end{bmatrix} \\
&+ \sum_{m=1}^{M}\sum_{\substack{t=1 \\ t\neq r}}^{R} \begin{bmatrix} Cov(E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_m}\right)(\hat{\pi}_m - \pi_m), E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{M+t}}\right)(\tilde{\pi}_{M+t} - \pi_{M+t})) \\ +Cov(E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{M+t}}\right)(\tilde{\pi}_{M+t} - \pi_{M+t}), E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_m}\right)(\hat{\pi}_m - \pi_m)) \end{bmatrix}.
\end{aligned} \tag{3.19}$$

We name the eight parts above (#1) through (#8) and calculate them separately. They are reported in Appendix II-A. After further simplification, we obtain the following:

$$V(\tilde{b}_{M+r}) = (G^{(M+r)\prime} I(\pi_{M+r}) G^{(M+r)})^{-1} V(\Lambda_{M+r}) (G^{(M+r)\prime} I(\pi_{M+r}) G^{(M+r)})^{-1}$$

$$= d_{1r} (G^{(M+r)\prime} I(\pi_{M+r}) G^{(M+r)})^{-1}$$

$$+ (G^{(M+r)\prime} I(\pi_{M+r}) G^{(M+r)})^{-1} G^{(M+r)\prime} \left[d_{2r} I(\pi_{M+r}) (X'X)^{-1} I(\pi_{M+r}) + P^{(M+r)} \right] \cdot$$

$$\cdot G^{(M+r)} (G^{(M+r)\prime} I(\pi_{M+r}) G^{(M+r)})^{-1}. \qquad (3.20)$$

Note that the asymptotic covariance matrices in equations (3.16) and (3.20) involve variances and covariances of reduced form error terms. They have to be estimated in stage 1 prior to estimating the corrected covariance matrices in stage 2. The variances and covariances from *M* equations with observed continuous dependent variables (y_j, j = 1, 2, ….., M) are estimated as

$$\hat{\sigma}_j^2 = \frac{\sum_{i=1}^{n} \hat{v}_{ji}^2}{n} \quad \text{and} \quad \hat{\sigma}_{jk} = \frac{\sum_{i=1}^{n} \hat{v}_{ji} \hat{v}_{ki}}{n}, \; j = k = 1, 2, \ldots\ldots, \text{M}; \quad j \neq k .$$

Following Amemiya (1978, p. 1200), the covariances between the reduced form error terms of observed y_j (j = *1, 2, ..., M*) and latent y_k^* (k = *M+1, ……, M+R*) are estimated as follows:

$$\hat{\sigma}_{jk} = \frac{\sum_{i=1}^{n} y_{ki} \hat{v}_{ji} f^{-1}(x_i' \tilde{\pi}_k)}{n},$$

where $j = 1, 2, ..., M$; $k = M+1, ..., M+R$ and x_i' is the i^{th} row of the matrix X. Finally, $\sigma_{kk'} = \rho_{kk'}$, the correlation between the error terms of latent y_k^* and $y_{k'}^*$ ($k = k' = M+1, ..., M+R,$ and $k \neq k'$), can be obtained by estimating the relevant parameter vectors $(\pi_k, \pi_{k'}; \rho_{kk'})$ by bivariate probit maximum likelihood method (Greene, 2012).

3.4 An Empirical Application

To see how these general formulas can be used to derive asymptotic variance-covariance matrices under specific situations, we present in this section a simultaneous equations model with one continuous and two binary dependent variables, and derive asymptotic variance-covariance matrices of the corresponding two-stage estimators. Mohanty (2009b) made an attempt to estimate such a simultaneous equations model, but did not report the corrected t-ratios of his estimates. We use the same data set and report two stage estimates of all three equations with corrected t-statistics. The derivation of these covariance matrices directly follows from our general formula and thus it shows the usefulness of this study for practitioners of econometrics.

3.4.1 Model

Mohanty's (2009b) model consists of three equations – log wage, happiness and positive attitude. Out of these three dependent variables, wage is continuous and the other two are binary. Mohanty assumes that these three variables affect each other and therefore are simultaneously related. Defining log wage $= y_1$, happiness $= y_2$ and positive attitude $= y_3$. Mohanty's model can then be rewritten in terms of our notations as follows:

$$
\begin{aligned}
y_1 &= Y^* \delta_1 + X_1 \beta_1 + \varepsilon_1, \\
y_2^* &= \gamma_2 y_1 + \delta_2 y_3^* + X_2 \beta_2 + \varepsilon_2, \\
y_3^* &= \gamma_3 y_1 + \delta_3 y_2^* + X_3 \beta_3 + \varepsilon_3, \\
y_2 &= 1, \quad if \;\; y_2^* > 0; \;\; = 0, \quad otherwise, \\
y_3 &= 1, \quad if \;\; y_3^* > 0; \;\; = 0, \quad otherwise,
\end{aligned}
$$

where each of y_1, y_2^*, y_3^* is an $n \times 1$ vector, $Y^* = (y_2^*, y_3^*)$, $\delta_1 = (\delta_{11}, \delta_{12})'$, X_1, X_2, X_3 are $n \times K_1, n \times K_2, n \times K_3$ matrices of exogenous variables respectively, and the error terms $\varepsilon' s$ follow the standard assumptions of classical regression models.

In the first stage, we estimate the following reduced model

$$
\begin{aligned}
y_1 &= X\pi_1 + v_1, \\
y_2^* &= X\pi_2 + v_2, \\
y_3^* &= X\pi_3 + v_3,
\end{aligned}
$$

where X is an $n \times K$ matrix of all distinct exogenous variables obtained from X_1, X_2, X_3, i.e., $X = [X_1 \bigcup X_2 \bigcup X_3]$, and π_1, π_2, π_3 are the reduced form parameters. Following the method and notations in section 3.2, we have the first stage estimators $\hat{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3$.

It follows that the second stage estimators are

$$\hat{b}_1 = (\hat{\delta}_1', \hat{\beta}_1')' = (\hat{H}_1' X' X \hat{H}_1)^{-1} \hat{H}_1' X' y_1,$$

where $\hat{H}_1 = [\tilde{\pi}_2, \tilde{\pi}_3, J_1]$, and $J_1 = [I_{K_1} \; 0_{K_1 \times (K - K_1)}]'$.

Then $\tilde{b}_{1+r} = (\tilde{\gamma}_{1+r} \; \tilde{\delta}_{1+r} \; \tilde{\beta}_{1+r}')'$, $r = 1, 2$ are obtained by maximizing $L(\hat{\pi}_1, \tilde{\pi}_{(-(1+r))}, b_{1+r})$.

As discussed in section 3.3, the corresponding variance-covariance matrices are as follows:

$$V(\hat{b}_1) = c_1 (\hat{H}_1' X' X \hat{H}_1)^{-1} + (\hat{H}_1' X' X \hat{H}_1)^{-1} \hat{H}_1' X' X V_{01} X' X \hat{H}_1 (\hat{H}_1' X' X \hat{H}_1)^{-1},$$

where $c_1 = \alpha_1^{(1)\prime} \Omega \alpha_2^{(1)} = (-1, 0, 0) \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & 1 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 2\delta_{11} \\ 2\delta_{12} \end{pmatrix} = \sigma_1^2 - 2(\delta_{11}\sigma_{12} + \delta_{12}\sigma_{13}),$

and $V_{01} = \delta_{11}^2 I^{-1}(\pi_2) + \delta_{12}^2 I^{-1}(\pi_3) + \delta_{11}\delta_{12} Q^{(2,3)}$,

where

$Q^{(2,3)} = I^{-1}(\pi_2) X' B^{(2,3)} X I^{-1}(\pi_3) + I^{-1}(\pi_3) X' B^{(2,3)} X I^{-1}(\pi_2)$,

and

$$B^{(2,3)} = \begin{pmatrix} \dfrac{(F_{2,3,1} - F_{2,1}F_{3,1})f_{2,1}f_{3,1}}{F_{2,1}(1-F_{2,1})F_{3,1}(1-F_{3,1})} & & 0 \\ & \ddots & \\ 0 & & \dfrac{(F_{2,3,n} - F_{2,n}F_{3,n})f_{2,n}f_{3,n}}{F_{2,n}(1-F_{2,n})F_{3,n}(1-F_{3,n})} \end{pmatrix}.$$

The variance-covariance matrices of $\tilde{b}_2$ and $\tilde{b}_3$ are as follows:

$$V(\tilde{b}_2) = d_{11}(G^{(2)\prime} I(\pi_2) G^{(2)})^{-1} + (G^{(2)\prime} I(\pi_2) G^{(2)})^{-1} G^{(2)\prime} \left[d_{21} I(\pi_2)(X'X)^{-1} I(\pi_2) + P^{(2)} \right] G^{(2)} (G^{(2)\prime} I(\pi_2) G^{(2)})^{-1},$$

where $G^{(2)} = [\pi_1, \pi_3, J_2]$, $J_2 = [I_{K_2}\, 0_{K_2 \times (K-K_2)}]'$, $d_{11} = 1 + \delta_2^2$,

$$d_{21} = \zeta_1^{(1)\prime} \Omega \zeta_2^{(1)} = (\gamma_2, 0, 0) \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & 1 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & 1 \end{pmatrix} \begin{pmatrix} \gamma_2 \\ -2 \\ 2\delta_2 \end{pmatrix} = \gamma_2^2 \sigma_1^2 - 2\gamma_2 \sigma_{12} + 2\gamma_2 \delta_2 \sigma_{13},$$

$P^{(2)} = -\delta_2 D^{(2,3)}$, where $D^{(2,3)} = X'B^{(2,3)} X I^{-1}(\pi_3) I(\pi_2) + I(\pi_2) I^{-1}(\pi_3) X' B^{(2,3)} X$,

and X is the matrix of all variables in X_1, X_2, and X_3 with X_2 variables placed in the beginning of the matrix, i.e., $X = [X_2 \cup X_1 \cup X_3]$. Similarly,

$$V(\tilde{b}_3) = d_{12}(G^{(3)\prime} I(\pi_3) G^{(3)})^{-1} + (G^{(3)\prime} I(\pi_3) G^{(3)})^{-1} G^{(3)\prime} \left[d_{22} I(\pi_3)(X'X)^{-1} I(\pi_3) + P^{(3)} \right] G^{(3)} (G^{(3)\prime} I(\pi_3) G^{(3)})^{-1},$$

where $G^{(3)} = [\pi_1, \pi_2, J_3]$, $J_3 = [I_{K_3}\, 0_{K_3 \times (K-K_3)}]'$, $d_{12} = 1 + \delta_3^2$,

$$d_{22} = \zeta_1^{(2)\prime} \Omega \zeta_2^{(2)} = (\gamma_3, 0, 0) \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & 1 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & 1 \end{pmatrix} \begin{pmatrix} \gamma_3 \\ 2\delta_3 \\ -2 \end{pmatrix} = \gamma_3^2 \sigma_1^2 + 2\gamma_3 \delta_3 \sigma_{12} - 2\gamma_3 \sigma_{13},$$

and $P^{(3)} = -\delta_3 D^{(3,2)}$, where $D^{(3,2)} = X'B^{(3,2)}XI^{-1}(\pi_2)I(\pi_3) + I(\pi_3)I^{-1}(\pi_2)X'B^{(3,2)}X$, $X = [X_3 \bigcup X_1 \bigcup X_2]$ and $B^{(3,2)} = B^{(2,3)}$ that has been defined earlier in this section.

Again, estimation of the corrected asymptotic variance-covariance matrices requires consistent estimation of the error structure from stage 1. We follow the methods given at the end of section 3.3 to achieve this estimation.

3.4.2 Data and Results

We draw a sample from the 1987 survey of the National Longitudinal Survey of Youth, 1979 (NLSY79). The three dependent variables as mentioned earlier are log hourly wage (*LOGWAGE*), "positive attitude to life" (*POSITIVE*) and "satisfied with life" (*HAPPY*). After eliminating missing observations from relevant variables, we obtained a sample of 7,379 employed workers.

The exogenous explanatory variables[2] that affect the worker's wage rate are years of schooling (*YEARSCHL*), intelligence measured by the Armed Force Qualifications Test (*AFQT*) score, experience (*EXP, EXPSQ*), tenure with the current employer (*TENURE*), full-time status (*FULLTIME*), marital status (*MARRIED*), health status (*HLTHPROB*), gender (*MALE*), race (*WHITE*), region of residence (*NOTHEAST, NOTHCENT, WEST*), location of residence (*CENTCITY, SMSBALNS*), job status (*UNION, GOVT*), workplace and employer sizes (*PLNTSIZE, FIRMSIZE*).

In addition, we assume that *HAPPY* and *POSITIVE* also affect the worker's wage rate positively.

Variables that are likely to affect the worker's happiness are age (*AGE, AGESQ*), marital status (*MARRIED*), health problems (*HLTHPROB*), fulltime employment status (*FULLTIME*), family income (*FAMINC*), family size (*FAMSIZE*), number of children (*CHILDNUM*), home ownership (*OWNHOUSE*), years of schooling (*YEARSCHL*), current school enrollment status (*ENROLL*), level of intelligence (*AFQT*), gender (*MALE*), race (*WHITE*) and location of residence (*CENTCITY, SMSBALNS*). In addition, we assume wage rate and positive attitude to affect the worker's happiness.

The variables that influence the worker's happiness are likely to affect his/her attitude. However, some variables, such as family size, number of children, full time employment status and current school enrollment status, which may affect happiness, are unlikely to affect the worker's attitude. These variables therefore are excluded from the positive attitude equation. Parental education and occupation (*MOTHGRAD, FATHGRAD, FATHMANG, FATHSALE, FATHCRFT, FATHOPRT, FATHSERV*), on the other hand, may have a genetic effect on the worker's attitude (Taubman, 1976) and hence are included in this equation. These variables are most unlikely to have any significant effect on the worker's current level of satisfaction and therefore are

excluded from the happiness equation. With these variable restrictions, all three equations are identified in this model. These variables are defined in Appendix II-B.

Following the two-step procedure, we estimate first the reduced forms of wage, happiness and positive attitude equations using *LIMDEP* [3] (Greene, 2007). These results are then used to compute the following:

$\hat{\sigma}_1^2 = 0.182158$, $\hat{\sigma}_{12} = 0.033829$, and $\hat{\sigma}_{13} = 0.0355847$.

With appropriate commands, *LIMDEP* routinely generates the following matrices and vectors to be used in the second stage estimation:

$Var(\hat{\pi}_2) = I^{-1}(\hat{\pi}_2)$, $Var(\hat{\pi}_3) = I^{-1}(\hat{\pi}_3)$, $\hat{y}_1 = X\hat{\pi}_1 = WAGEHAT$,

$\tilde{y}_2{}^* = X\tilde{\pi}_2 = HAPPYHAT$, and $\tilde{y}_3{}^* = X\tilde{\pi}_3 = POSHAT$.

With these estimated variables, the structural wage, happiness and attitude equations were estimated in stage 2. These structural coefficients are reported in Table 3.1. Using these structural coefficients, we obtain the following scalars for corrected covariance matrices:

$c_1 = 0.138557513$, $d_{11} = 1.446556499$, $d_{21} = 0.008655284$, $d_{12} = 1.581227579$, $d_{22} = 0.003470492$. Also we get $\rho_{23} = 0.8526$ which is required to generate the diagonal matrix $B^{(2,3)}$. Finally, these results are used to compute the asymptotic

variance-covariance matrices of the three structural coefficient vectors and their t-ratios. These corrected t-statistics are reported in Table 3.1. [4]

Results in Table 3.1 indicate that the worker's wage is positively correlated with his/her positive attitude, but not necessarily with happiness. Interestingly, happiness is found to be positively correlated with attitude and negatively correlated with wage income (Mohanty, 2009b), confirming thus the Easterlin paradox (Easterlin, 1974) that money does not necessarily buy happiness. As expected, an individual's positive attitude is found to be positively correlated with both the objective factor wage and the subjective factor happiness. Interestingly, both the psychological variables attitude and happiness are positively related to each other which clearly suggests that happiness affects wage indirectly through its direct effect on the worker's attitude (Mohanty, 2009b). We summarize the relationship of the three dependent variables in the following chart.

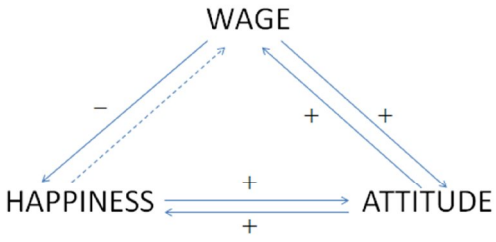


Figure 3.1: Relationship of dependent variables where solid(dash) arrow represents (in)direct effect and +(-) sign shows positive(negative) effect

3.5 Conclusion

This study presents general formulas for asymptotic variance-covariance matrices of two-stage estimators for a simultaneous equations system with a mixture of binary and continuous dependent variables. We demonstrate how our formulas can be used to write covariance matrices of several two-stage estimators in a straight--forward manner. We also examine an empirical application to illustrate our procedure using the econometric software *LIMDEP*. For now, *LIMDEP* only has the simultaneous probit model with two dependent variables with corrected asymptotic variance-covariance matrices built in its package. After this study, *LIMDEP* may attempt to build in some more complex simultaneous equations system with corrected asymptotic variance-covariance matrices. Clearly, these results would be of considerable help to practitioners of econometrics. In the future, we will also try to build in other standard statistical software packages, say, *R*, the corrected variance-covariance matrices of two-stage estimators for such general simultaneous equation models with any number of continuous and binary dependent variables.

Table 3.1 Two-Stage Estimates of Wage, Happiness and Attitude Equations[a]

Variable	Wage	Happiness	Attitude
Constant	1.5365**	1.9322	-1.1283
	(11.088)	(1.133)	(0.682)
WAGEHAT	_____	-0.1697*	0.2019**
		(1.669)	(2.204)
HAPPYHAT	0.1076	_____	0.7624**
	(0.872)		(5.020)
POSTVHAT	0.5103**	0.6682**	_____
	(4.062)	(5.444)	
YEARSCHL	-0.0108	0.0100	0.0169*
	(1.432)	(0.987)	(1.680)
AFQT	0.0014**	0.0017*	-0.0005
	(2.423)	(2.445)	(0.664)
EXP	0.0496**	_____	_____
	(3.032)		
EXPSQ	-0.0011	_____	_____
	(0.825)		
TENURE	0.0255**	_____	_____
	(5.453)		
FULLTIME	0.0372	_____	_____
	(1.469)		
GOVT	0.0424	_____	_____
	(1.337)		
UNION	0.0779**	_____	_____
	(3.032)		
PLNTSIZE	0.1329**	_____	_____
	(4.329)		
FIRMSIZE	0.0084	_____	_____
	(0.359)		
AGE	_____	-0.1729	0.0539
		(1.302)	(0.410)
AGESQ	_____	0.0033	-0.0011
		(1.303)	(0.433)
MARRIED	0.0219	0.0582*	-0.0281
	(1.010)	(1.857)	(1.005)
MALE	0.1039**	-0.0255	0.0747**

	(3.993)	(0.675)	(2.241)
WHITE	0.1650**	0.0404	-0.1786**
	(4.839)	(0.848)	(4.879)
CENTCITY	0.0174	-0.0474	0.1095**
	(0.545)	(1.177)	(3.208)
SMSBALNS	0.0783**	0.0373	-0.0066
	(2.650)	(0.962)	(0.175)
NOTHEAST	0.1334**	_____	_____
	(4.662)		
NOTHCENT	0.0049	_____	_____
	(0.177)		
WEST	0.0682**	_____	_____
	(2.425)		
HLTHPROB	-0.0421	-0.0900	0.0584
	(0.699)	(1.144)	(0.775)
FAMINC	_____	0.0012	-0.0003
		(1.224)	(0.294)
OWNHOUSE	_____	0.0828**	-0.0562
		(2.495)	(1.643)
FULLTIME	_____	0.0486	_____
		(1.499)	
ENROLL	_____	0.1023**	_____
		(2.158)	
FAMSIZE	_____	0.0030	_____
		(0.300)	
CHILDNUM	_____	-0.0359**	_____
		(1.988)	
MOTHGRAD	_____	_____	0.0057
			(1.247)
FATHGRAD	_____	_____	0.0073**
			(2.248)
FATHMANG	_____	_____	-0.0907**
			(2.306)
FATHSALE	_____	_____	0.0253
			(0.463)
FATHCRFT	_____	_____	-0.0186
			(0.522)
FATHOPRT	_____	_____	-0.0244
			(0.613)
FATHSERV	_____	_____	0.1220**
			(2.178)

Sample Size	7,357	7,379	7,379
Adjusted R^2	0.2963	_____	_____
Log Likelihood	-4,375.543	-4,574.168	-4,861.326
Chi-Squared (df)	2,613.58 (21)	243.03 (18)	384.94 (21)

a The quantities in parentheses are the absolute t-ratios.
** (*) Significant at 5 (10) percent level.

Note

1. The assumption of unit variance of v_{M+r} is standard in probit analysis. This avoids estimating regression parameters to a scalar proportion since the variance of v_{M+r} is unidentifiable.

2. The exogenous explanatory variables in each equation are selected by experience or by following previous models, like those used in Mohanty (2009b).

3. The name *LIMDEP* is derived from LIMited DEPendent models. This econometric software package was originally written by William H. Greene and is now in wide use by econometricians.

4. For comparison use, we also provide the results of this empirical application with uncorrected standard errors in Appendix II-C.

Bibliography

[1] Amemiya, Takeshi, "The Estimation of a Simultaneous Equation Generalized Probit Model," *Econometrica*, 46 (1978): 1193-1205.

[2] Amemiya, Takeshi, "The Estimation of a Simultaneous Equation Tobit Model," *International Economic Review*, 20 (1979): 169-81.

[3] Bierens, Herman J., "Estimation a System of Simultaneous Probit Models by Nonlinear GMM," *EasyReg International Guided Tour.*

[4] Blanchard, L. J., "The Production and Inventory Behavior of the American Automobile Industry," *Journal of Political Economy*, 91 (1985): 365-400.

[5] Brundy, J., and D. Jorgenson, "Consistent and Efficient Estimation of Systems of Simultaneous Equations by Means of Instrumental Variables," *Review of Economics and Statistics*, 53 (1971): 207-24.

[6] Easterlin, Richard A., "Does Economic Growth Improve the Human Lot? Some Empirical Evidence," In P. A. David and M. W. Reder (eds), *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, Academic Press, New York, and London (1974): 89-125.

[7] Greene, William H., *Econometric Analysis*, 7th Edition, Prentice Hall, New York (2012).

[8] Greene, William H., *Limdep, Version 9.0*, Econometric Software Inc., 15 Gloria Place, Plainview, New York (2007).

[9] Heckman, James, "Sample Selection Bias as a Specification Error," *Econometrica*, 47 (1979): 153-61.

[10] Heywood, John S., and Madhu S. Mohanty, "Estimation of the Federal Job Queue in the Presence of Endogenous Union Queue," *Economica*, 62 (1995): 479-93.

[11] Lee, Lung Fee, "Estimation of Limited Dependent Variable Models by Two-Stage Methods," *Ph. D. Dissertation*, University of Rochester (1976).

[12] Lee, Lung Fee, G. S. Maddala, and R. P. Prost, "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity," *Econometrica*, 48 (1980): 401-503.

[13] Lydon, Reamonn, and Arnaud Chevalier, "Estimates of Wages on Job Satisfaction," *Center for Economic Performance*, London School of Economics and Political Science (May 2002).

[14] Maddala, G. S., *Limited Dependent and Qualitative Variables in Econometrics,* Cambridge University Press, Cambridge (1983).

[15] Mohanty, Madhu S., "Asymptotic Properties of the Two-Stage Bivariate Probit Estimator in the Presence of Partial Observability," *Journal of Quantitative Economics*, 11 (1995): 101-25.

[16] Mohanty, Madhu S., "The Role of the "Desire for Union Status" in the Decision to Enter the Local Government Job Queue," *Applied Economics*, 30 (1998): 1369-78.

[17] Mohanty, Madhu S., **"**Effects of Positive Attitude on Earnings: Evidence from the US Longitudinal Data," *Journal of Socio-Economics*, 38 (2009a): 357-71.

[18] Mohanty, Madhu S., "Effects of Positive Attitude on Happiness and Wage: Evidence from the US Data," *Journal of Economic Psychology*, 30 (2009b): 884-97.

[19] Mohanty, Madhu S., and Aman Ullah, "Direct and Indirect Effects of Happiness on Wage: A Simultaneous Equations Approach," *Journal of Socio-Economics*, 41 (2012): 143-52.

[20] Murphy, Kevin M., and Robert H. Topel, "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics*, 3 (1985): 88-97.

[21] Tunali, Insan, "A General Structures for Models of Double Selection and an Application to a Joint Migration/Earning Process with Remigration," *Research in Labor Economics*, 8 (1986): 235-82.

[22] Taubman, P., "Earnings, Education, Genetics, and Environment," *Journal of Human Resources*, 11 (1976): 447-61.

[23] Windmeijer, F.A.G., Santos Silva, J.M.C., "Endogeneity in Count Data Models: An Application to Demand for Health Care," *Journal of Applied Econometrics*, 12 (1997): 281-94.

[24] Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, 2nd edition (2010).

Appendix II-A Technical Support

The different terms of equation (3.15) are derived as follows:

$(*1) = V(v_m) = \sigma_m^2 I$.

$$(*2) = \sum_{\substack{s=1 \\ s\neq m}}^{M} V(\gamma_{ms} X(X'X)^{-1} X' v_s) = \sum_{\substack{s=1 \\ s\neq m}}^{M} \gamma_{ms}^2 \sigma_s^2 X(X'X)^{-1} X' .$$

$$(*3) = \sum_{r=1}^{R} \delta_{mr}^2 X I^{-1}(\pi_{M+r}) X' .$$

Since $Cov(v_m, \gamma_{ms} X(\hat{\pi}_s - \pi_s)) = Cov(\gamma_{ms} X(\hat{\pi}_s - \pi_s), v_m) = \gamma_{ms}\sigma_{ms} X(X'X)^{-1} X'$,

$$(*4) = -2\sum_{\substack{s=1 \\ s\neq m}}^{M} \gamma_{ms}\sigma_{ms} X(X'X)^{-1} X' .$$

To get (*5), we need the expression for $I(\pi_{M+r})$ and $E(v_m S'(\pi_{M+r}))$.

From equation (3.8), we have

$$\log L(\pi_{M+r}) = \sum_{i=1}^{n} \left[y_{M+r,i} \log F_{M+r,i} + (1 - y_{M+r,i}) \log(1 - F_{M+r,i}) \right],$$

where $F_{M+r,i} = P(v_{M+r} \le x_i' \pi_{M+r}) = \Phi(x_i' \pi_{M+r})$ and Φ is the standard normal cdf.

It follows that

$$S(\pi_{M+r}) = \sum_{i=1}^{n} \frac{y_{M+r,i} - F_{M+r,i}}{F_{M+r,i}(1 - F_{M+r,i})} f_{M+r,i} x_i \text{ , and}$$

$$\begin{aligned} I(\pi_{M+r}) &= \sum_{i=1}^{n} \frac{f_{M+r,i}^2 x_i x_i'}{F_{M+r,i}\left(1 - F_{M+r,i}\right)} \\ &= (x_1, \cdots, x_n) diag\left(\frac{f_{M+r,i}^2}{F_{M+r,i}\left(1 - F_{M+r,i}\right)} \right) \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix} \\ &\equiv X' A^{(M+r)} X. \end{aligned}$$

With some further derivation, we get $E(v_m S'(\pi_{M+r})) = \sigma_{m,M+r} A^{(M+r)} X$.

Hence

$$Cov(v_m, \delta_{mr} X(\tilde{\pi}_{M+r} - \pi_{M+r})) = \delta_{mr} E(v_m S'(\pi_{M+r})) I^{-1}(\pi_{M+r}) X'$$

$$= \delta_{mr} \sigma_{m,M+r} A^{(M+r)} X (X' A^{(M+r)} X)^{-1} X',$$

and $Cov(\delta_{mr} X(\tilde{\pi}_{M+r} - \pi_{M+r}), v_m) = \delta_{mr} \sigma_{m,M+r} X (X' A^{(M+r)} X)^{-1} X' A^{(M+r)}$, providing that

$$(*5) = -\sum_{r=1}^{R} \delta_{mr} \sigma_{m,M+r} \left[A^{(M+r)} X (X' A^{(M+r)} X)^{-1} X' + X (X' A^{(M+r)} X)^{-1} X' A^{(M+r)} \right].$$

$$(*6) = \sum_{\substack{s=1 \\ s \neq m}}^{M} \sum_{\substack{t=1 \\ t \neq m \\ t \neq s}}^{M} \gamma_{ms} \gamma_{mt} \sigma_{st} X (X'X)^{-1} X'.$$

The expression of $E[S(\pi_{M+r_1}) S'(\pi_{M+r_2})]$ is required to compute (*7). We have

$$E[S(\pi_{M+r_1}) S'(\pi_{M+r_2})] = \sum_{i=1}^{n} \frac{E[(y_{M+r_1,i} - F_{M+r_1,i})(y_{M+r_2,i} - F_{M+r_2,i})] f_{M+r_1,i} f_{M+r_2,i}}{F_{M+r_1,i}(1 - F_{M+r_1,i}) F_{M+r_2,i}(1 - F_{M+r_2,i})} x_i x_i'$$

$$= \sum_{i=1}^{n} \frac{(F_{M+r_1,M+r_2,i} - F_{M+r_1,i} F_{M+r_2,i}) f_{M+r_1,i} f_{M+r_2,i}}{F_{M+r_1,i}(1 - F_{M+r_1,i}) F_{M+r_2,i}(1 - F_{M+r_2,i})} x_i x_i'$$

$$\equiv X' B^{(M+r_1, M+r_2)} X,$$

where $F_{M+r_1,M+r_2,i}$ is the joint cdf of bivariate normal $(v_{M+r_1,i}, v_{M+r_2,i})$, and

$$B^{(M+r_1,M+r_2)} = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_n \end{pmatrix},$$

where $B_i = \dfrac{(F_{M+r_1,M+r_2,i} - F_{M+r_1,i} F_{M+r_2,i}) f_{M+r_1,i} f_{M+r_2,i}}{F_{M+r_1,i}(1 - F_{M+r_1,i}) F_{M+r_2,i}(1 - F_{M+r_2,i})}$, $i = 1, \cdots, n$.

From this,

$$(*7) = \frac{1}{2}\sum_{r_1=1}^{R}\sum_{\substack{r_2=1\\ r_2\neq r_1}}^{R}\delta_{mr_1}\delta_{mr_2}X\left\{\begin{matrix}(X'A^{(M+r_1)}X)^{-1}E[S(\pi_{M+r_1})S'(\pi_{M+r_2})](X'A^{(M+r_2)}X)^{-1}+\\ (X'A^{(M+r_2)}X)^{-1}E[S(\pi_{M+r_1})S'(\pi_{M+r_2})](X'A^{(M+r_1)}X)^{-1}\end{matrix}\right\}X'$$

$$= \frac{1}{2}\sum_{r_1=1}^{R}\sum_{\substack{r_2=1\\ r_2\neq r_1}}^{R}\delta_{mr_1}\delta_{mr_2}X\left\{\begin{matrix}(X'A^{(M+r_1)}X)^{-1}X'B^{(M+r_1,M+r_2)}X(X'A^{(M+r_2)}X)^{-1}+\\ (X'A^{(M+r_2)}X)^{-1}X'B^{(M+r_1,M+r_2)}X(X'A^{(M+r_1)}X)^{-1}\end{matrix}\right\}X'$$

$$\equiv \frac{1}{2}\sum_{r_1=1}^{R}\sum_{\substack{r_2=1\\ r_2\neq r_1}}^{R}\delta_{mr_1}\delta_{mr_2}XQ^{(M+r_1,M+r_2)}X'.$$

Finally,

$$(*8) = 2\sum_{\substack{s=1\\ s\neq m}}^{M}\sum_{r=1}^{R}\gamma_{ms}\delta_{mr}\sigma_{s,M+r}X(X'X)^{-1}X'.$$

By adding up (*1) through (*8) and pre-multiplying X' and post-multiplying X, we obtain

$$X'V(w_m)X = c_m X'X + X'XV_{0m}X'X,$$

where

$$c_m = \sigma_m^2 + \sum_{\substack{s=1\\ s\neq m}}^{M}\left(\gamma_{ms}^2\sigma_s^2 - 2\gamma_{ms}\sigma_{ms} + \sum_{\substack{t=1\\ t\neq m\\ t\neq s}}^{M}\gamma_{ms}\gamma_{mt}\sigma_{st} + 2\sum_{r=1}^{R}\gamma_{ms}\delta_{mr}\sigma_{s,M+r}\right) - 2\sum_{r=1}^{R}\delta_{mr}\sigma_{m,M+r}$$

$$\equiv \alpha_1^{(m)\prime}\Omega\alpha_2^{(m)},$$

in which $\Omega = \begin{pmatrix}\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1,M+R}\\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2,M+R}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{1,M+R} & \sigma_{2,M+R} & \cdots & \sigma_{M+R}^2\end{pmatrix}$, $\alpha_1^{(m)} = \begin{pmatrix}\gamma_m^*\\ 0_{R\times 1}\end{pmatrix}$ and $\alpha_2^{(m)} = \begin{pmatrix}\gamma_m^*\\ 2\delta_m\end{pmatrix}$,

where

$\gamma_m^* = (\gamma_{m1}, \cdots, \gamma_{m,m-1}, -1, \gamma_{m,m+1}, \cdots, \gamma_{mM})'$.

Also

$$V_{0m} = \sum_{r=1}^{R} \delta_{mr}^2 I^{-1}(\pi_{M+r}) + \frac{1}{2} \sum_{r_1=1}^{R} \sum_{\substack{r_2=1 \\ r_2 \neq r_1}}^{R} \delta_{mr_1} \delta_{mr_2} Q^{(M+r_1, M+r_2)}$$

$$\equiv (\delta_m' \otimes I_K) \Delta (\delta_m \otimes I_K) + \frac{1}{2} (\delta_m' \otimes 1_{R-1}' \otimes I_K) \Theta (\delta_m^{(-)} \otimes I_K),$$

in which K is the second dimension of matrix X, and Δ is a block-diagonal matrix with $I^{-1}(\pi_{M+r})$ being its r^{th} diagonal.

$\Theta \equiv blockdiag(Q^{(M+1,M+2)}, Q^{(M+1,M+3)}, \cdots Q^{(M+1,M+R)}, Q^{(M+2,M+1)}, Q^{(M+2,M+3)}, \cdots Q^{(M+R,M+R-1)})$,

and $\delta_m^{(-)} = (\delta_m^{(-1)\prime}, \cdots, \delta_m^{(-R)\prime})'$ where $\delta_m^{(-r)} = (\delta_{m1}, \cdots, \delta_{m,r-1}, \delta_{m,r+1}, \cdots, \delta_{mR})'$.

Entering this expression of $X'V(w_m)X$ into equation (3.13), we obtain the following neat formula of the variance-covariance matrix,

$$V(\hat{b}_m) = (\hat{H}_m' X'X \hat{H}_m)^{-1} \hat{H}_m' X'V(w_m) X \hat{H}_m (\hat{H}_m' X'X \hat{H}_m)^{-1}$$

$$= c_m (\hat{H}_m' X'X \hat{H}_m)^{-1} + (\hat{H}_m' X'X \hat{H}_m)^{-1} \hat{H}_m' X'X V_{0m} X'X \hat{H}_m (\hat{H}_m' X'X \hat{H}_m)^{-1}.$$

This matrix is reported as equation (3.16) in the text.

The eight different terms of equation (3.19) are derived in this subsection. Before calculating them, however, we obtain from equations (3.11) and (3.12) the following items:

$$\frac{\partial \log L}{\partial b_{M+r}} = G^{(M+r)\prime}\frac{\partial \log L}{\partial \pi_{M+r}} = G^{(M+r)\prime}S(\pi_{M+r})$$

$$E\left(\frac{-\partial^2 \log L}{\partial b_{M+r}\partial b'_{M+r}}\right) = G^{(M+r)\prime}(X'A^{(M+r)}X)G^{(M+r)} = G^{(M+r)\prime}I(\pi_{M+r})G^{(M+r)}$$

$$E\left(\frac{\partial \log L}{\partial b_{M+r}}v'_m\right) = \sigma_{m,M+r}G^{(M+r)\prime}X'A^{(M+r)}, \qquad m = 1,\cdots,M$$

$$E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_m}\right) = -\gamma_{M+r,m}G^{(M+r)\prime}I(\pi_{M+r}), \qquad m = 1,\cdots,M$$

$$E\left(\frac{\partial^2 \log L}{\partial b_{M+r}\partial \pi'_{M+t}}\right) = -\delta_{M+r,t}G^{(M+r)\prime}I(\pi_{M+r}), \qquad t = 1,\cdots r-1, r+1,\cdots R$$

$$E\left(\frac{\partial \log L}{\partial b_{M+r}}(\tilde{\pi}_{M+t} - \pi_{M+t})'\right) = E\left(\frac{\partial \log L}{\partial b_{M+r}}S'(\pi_{M+t})\right)I^{-1}(\pi_{M+t})$$

$$= G^{(M+r)\prime}X'B^{(M+r,M+t)}XI^{-1}(\pi_{M+t}).$$

Using these, we get

$$(\#1) = V(\frac{\partial \log L}{\partial b_{M+r}}) = G^{(M+r)\prime}I(\pi_{M+r})G^{(M+r)}.$$

$$(\#2) = \sum_{m=1}^{M}\gamma_{M+r,m}^2\sigma_m^2 G^{(M+r)\prime}I(\pi_{M+r})(X'X)^{-1}I(\pi_{M+r})G^{(M+r)}.$$

$$(\#3) = \sum_{\substack{t=1\\ t\neq r}}^{R}\delta_{M+r,t}^2 G^{(M+r)\prime}I(\pi_{M+r})G^{(M+r)}.$$

$$(\#4) = -2\sum_{m=1}^{M}\gamma_{M+r,m}\sigma_{m,M+r}G^{(M+r)\prime}I(\pi_{M+r})(X'X)^{-1}I(\pi_{M+r})G^{(M+r)}.$$

$$(\#5) = -\sum_{\substack{t=1 \\ t\neq r}}^{R} \delta_{M+r,t} G^{(M+r)\prime} \begin{bmatrix} X'B^{(M+r,M+t)} XI^{-1}(\pi_{M+t}) I(\pi_{M+r}) \\ +I(\pi_{M+r}) I^{-1}(\pi_{M+t}) X'B^{(M+r,M+t)} X \end{bmatrix} G^{(M+r)}$$

$$\equiv -\sum_{\substack{t=1 \\ t\neq r}}^{R} \delta_{M+r,t} G^{(M+r)\prime} D^{(M+r,M+t)} G^{(M+r)}.$$

$$(\#6) = \sum_{m_1=1}^{M} \sum_{\substack{m_2=1 \\ m_2\neq m_1}}^{M} \gamma_{M+r,m_1} \gamma_{M+r,m_2} \sigma_{m_1,m_2} G^{(M+r)\prime} I(\pi_{M+r}) (X'X)^{-1} I(\pi_{M+r}) G^{(M+r)}.$$

$$(\#7) = \frac{1}{2} \sum_{\substack{s=1 \\ s\neq r}}^{R} \sum_{\substack{t=1 \\ t\neq r \\ t\neq s}}^{R} \delta_{M+r,s} \delta_{M+r,t} G^{(M+r)\prime} I(\pi_{M+r}) Q^{(M+s,M+t)} I(\pi_{M+r}) G^{(M+r)}.$$

$$(\#8) = 2 \sum_{m=1}^{M} \sum_{\substack{t=1 \\ t\neq r}}^{R} \gamma_{M+r,m} \delta_{M+r,t} \sigma_{m,M+t} G^{(M+r)\prime} I(\pi_{M+r}) (X'X)^{-1} I(\pi_{M+r}) G^{(M+r)}.$$

By adding up the eight terms above, we have

$$\begin{aligned} V(\Lambda_{M+r}) = {} & d_{1r} G^{(M+r)\prime} I(\pi_{M+r}) G^{(M+r)} + d_{2r} G^{(M+r)\prime} I(\pi_{M+r}) (X'X)^{-1} I(\pi_{M+r}) G^{(M+r)} \\ & - \sum_{\substack{t=1 \\ t\neq r}}^{R} \delta_{M+r,t} G^{(M+r)\prime} D^{(M+r,M+t)} G^{(M+r)} \\ & + \frac{1}{2} \sum_{\substack{s=1 \\ s\neq r}}^{R} \sum_{\substack{t=1 \\ t\neq r \\ t\neq s}}^{R} \delta_{M+r,s} \delta_{M+r,t} G^{(M+r)\prime} I(\pi_{M+r}) Q^{(M+s,M+t)} I(\pi_{M+r}) G^{(M+r)}, \end{aligned}$$

where $d_{1r} = 1 + \sum_{\substack{t=1 \\ t\neq r}}^{R} \delta_{M+r,t}^2 = \delta_{M+r}^{*\prime} \delta_{M+r}^{*}$, where we define

$$\delta_{M+r}^{*} = (\delta_{M+r,1}, \cdots, \delta_{M+r,r-1}, -1, \delta_{M+r,r+1}, \cdots, \delta_{M+r,R})'.$$

$$d_{2r} = \sum_{m=1}^{M} (\gamma_{M+r,m}^2 \sigma_m^2 - 2\gamma_{M+r,m} \sigma_{m,M+r} + 2 \sum_{\substack{t=1 \\ t\neq r}}^{R} \gamma_{M+r,m} \delta_{M+r,t} \sigma_{m,M+t}) + \sum_{m_1=1}^{M} \sum_{\substack{m_2=1 \\ m_2\neq m_1}}^{M} \gamma_{M+r,m_1} \gamma_{M+r,m_2} \sigma_{m_1 m_2}$$

$$\equiv \zeta_1^{(r)\prime} \Omega \zeta_2^{(r)},$$

where $\zeta_1^{(r)} = \begin{pmatrix} \gamma_{M+r} \\ 0_{R\times 1} \end{pmatrix}$, and $\zeta_2^{(r)} = \begin{pmatrix} \gamma_{M+r} \\ 2\delta_{M+r}^{*} \end{pmatrix}$.

Also we define a matrix $P^{(M+r)}$ that

$$P^{(M+r)} = \frac{1}{2}\sum_{\substack{s=1\\ s\neq r}}^{R}\sum_{\substack{t=1\\ t\neq r\\ t\neq s}}^{R}\delta_{M+r,s}\delta_{M+r,t}I(\pi_{M+r})Q^{(M+s,M+t)}I(\pi_{M+r}) - \sum_{\substack{t=1\\ t\neq r}}^{R}\delta_{M+r,t}D^{(M+r,M+t)}$$

$$\equiv \frac{1}{2}I(\pi_{M+r})(\delta'_{M+r}\otimes 1'_{R-2}\otimes I_K)\Theta^{*}(\delta^{(-)}_{M+r}\otimes I_K)I(\pi_{M+r}) - (\delta'_{M+r}\otimes I_K)\Psi^{(-r)}(1_{R-1}\otimes I_K),$$

where Θ^{*} is a block-diagonal matrix with its diagonals equal to $Q^{(M+s,M+t)}$ in the order that first keep $s=1$ (if $r\neq 1$) and let $t=2,\cdots,r-1,r+1,\cdots,R$, then $s=2$ and $t=1,3,\cdots,r-1,r+1,\cdots,R$,etc. And $\delta^{(-)}_{M+r}=(\delta^{(-1)'}_{M+r},\cdots,\delta^{(-r+1)'}_{M+r},\delta^{(-r-1)'}_{M+r},\cdots,\delta^{(-R)'}_{M+r})'$, where $\delta^{(-i)}_{M+r}$ is δ_{M+r} with the element $\delta_{M+r,i}$ deleted. Also $\Psi^{(-r)}$ is a block-diagonal matrix with $D^{(M+r,M+t)}(t\neq r)$ being its t^{th} diagonal.

With these notations, $V(\tilde{b}_{M+r})$ in equation (3.6) can be written in a neat form as

$$\begin{aligned}
V(\tilde{b}_{M+r}) &= (G^{(M+r)'}I(\pi_{M+r})G^{(M+r)})^{-1}V(\Lambda_{M+r})(G^{(M+r)'}I(\pi_{M+r})G^{(M+r)})^{-1}\\
&= d_{1r}(G^{(M+r)'}I(\pi_{M+r})G^{(M+r)})^{-1}\\
&\quad + (G^{(M+r)'}I(\pi_{M+r})G^{(M+r)})^{-1}G^{(M+r)'}\left[d_{2r}I(\pi_{M+r})(X'X)^{-1}I(\pi_{M+r}) + P^{(M+r)}\right]\cdot\\
&\quad \cdot G^{(M+r)}(G^{(M+r)'}I(\pi_{M+r})G^{(M+r)})^{-1}.
\end{aligned}$$

Appendix II-B Variable Definition of Empirical Data

Variable	Definition	Mean	Std. Dev
Log Wage	= log hourly wage	1.9545	0.5227
YEARSCHL	= Completed years of schooling	12.8245	2.4854
AFQT	= Armed Force Qualifications Test score	41.5170	29.4297
EXP	= Years of work experience	5.9745	2.4757
EXPSQ	= Experience squared	41.8233	29.9411
AGE	= Worker's age in years	25.8914	2.2538
AGESQ	= Age squared	675.446	117.168
HLTHPROB	= 1, if the worker has some health problem	0.0283	0.1659
MARRIED	= 1, if the worker is married with spouse present	0.4464	0.4972
NOTHEAST	= 1, if the worker lives in North-East	0.1888	0.3914
NOTHCENT	= 1, if the worker lives in North-Central	0.2262	0.4184
WEST	= 1, if the worker lives in West	0.1882	0.3909
CENTCITY	= 1, if the worker lives in central city	0.4735	0.4993
SMSBALNS	= 1, if the worker lives in the balance of SMSA	0.2930	0.4552
WHITE	= 1, if the worker is white	0.6856	0.4643
MALE	= 1, if the worker is male	0.5273	0.4993
TENURE	= Years of work experience with the current employer	2.5258	2.4990
FULLTIME	= 1, if the worker works 40 hours a week	0.7772	0.4162
GOVT	= 1, if the worker has a government job	0.1163	0.3206
UNION	= 1, if the worker has a union job	0.2134	0.4098
PLNTSIZE	= 1, if the number of workers in the current location is at least 500	0.1564	0.3632
FIRMSIZE	= 1, if the number of workers in all locations of the firm is 1,000 or more	0.3289	0.4698
HAPPY	= 1, if the worker very strongly claims that he/she is satisfied with life	0.3331	0.4713
POSITIVE	= 1, if the worker very strongly claims that he/she has a positive attitude to life	0.4358	0.4959
ENROLL	= 1, if the worker is currently enrolled in school	0.0741	0.2620
FAMINC	= Income of other family members (in $1,000)	24.193	21.161
FAMSIZE	= Number of people in the family	2.9165	1.6716
CHILDNUM	= Number of own children	0.6533	0.9559

OWNHOUSE	= 1, if the worker owns a house	0.2744	0.4463
MOTHGRAD	= Years of schooling completed by the mother	10.418	3.9741
FATHGRAD	= Years of schooling completed by The father	9.7456	5.1957
FATHMANG	= 1, if father had a managerial or professional job in 1980	0.1965	0.3974
FATHSALE	= 1, if father had a sales job in 1980	0.0680	0.2518
FATHCRFT	= 1, if father had a craft job in 1980	0.1794	0.3837
FATHOPRT	= 1, if father had an operative job in 1980	0.1379	0.3449
FATHSERV	= 1, if father had a service job in 1980	0.0569	0.2317
Sample Size	7,379		

Appendix II-C Results of Empirical Application with Uncorrected Test Statistics

Variable	Wage	Happiness	Attitude
Constant	1.5365**	1.9322	-1.1283
	(21.256)	(0.912)	(0.536)
WAGEHAT	_____	-0.1697	0.2019*
		(1.355)	(1.743)
HAPPYHAT	0.1076*	_____	0.7624**
	(1.658)		(3.972)
POSTVHAT	0.5103**	0.6682**	_____
	(7.755)	(4.427)	
YEARSCHL	-0.0108**	0.0100	0.0169
	(2.756)	(0.806)	(1.342)
AFQT	0.0014**	0.0017**	-0.0005
	(4.609)	(1.984)	(0.526)
EXP	0.0496**	_____	_____
	(5.810)		
EXPSQ	-0.0011	_____	_____
	(1.600)		
TENURE	0.0255**	_____	_____
	(10.417)		
FULLTIME	0.0372**	_____	_____
	(2.803)		
GOVT	0.0424**	_____	_____
	(2.546)		
UNION	0.0779**	_____	_____
	(5.758)		
PLNTSIZE	0.1329**	_____	_____
	(8.222)		
FIRMSIZE	0.0084	_____	_____
	(0.682)		
AGE	_____	-0.1729	0.0539
		(1.047)	(0.323)
AGESQ	_____	0.0033	-0.0011
		(1.047)	(0.341)
MARRIED	0.0219*	0.0582	-0.0281
	(1.927)	(1.499)	(0.794)
MALE	0.1039**	-0.0255	0.0747*
	(7.624)	(0.547)	(1.767)

WHITE	0.1650**	0.0404	-0.1786**
	(9.226)	(0.690)	(3.874)
CENTCITY	0.0174	-0.0474	0.1095**
	(1.043)	(0.951)	(2.516)
SMSBALNS	0.0783**	0.0373	-0.0066
	(5.085)	(0.770)	(0.137)
NOTHEAST	0.1334**	_____	_____
	(8.865)		
NOTHCENT	0.0049	_____	_____
	(0.339)		
WEST	0.0682**	_____	_____
	(4.619)		
HLTHPROB	-0.0421	-0.0900	0.0584
	(1.343)	(0.942)	(0.615)
FAMINC	_____	0.0012	-0.0003
		(0.992)	(0.233)
OWNHOUSE	_____	0.0828**	-0.0562
		(1.999)	(1.298)
FULLTIME	_____	0.0486	_____
		(1.216)	
ENROLL	_____	0.1023*	_____
		(1.705)	
FAMSIZE	_____	0.0030	_____
		(0.244)	
CHILDNUM	_____	-0.0359	_____
		(1.619)	
MOTHGRAD	_____	_____	0.0057
			(0.986)
FATHGRAD	_____	_____	0.0073*
			(1.780)
FATHMANG	_____	_____	-0.0907*
			(1.824)
FATHSALE	_____	_____	0.0253
			(0.367)
FATHCRFT	_____	_____	-0.0186
			(0.412)
FATHOPRT	_____	_____	-0.0244
			(0.483)
FATHSERV	_____	_____	0.1220*
			(1.755)

** (*) Significant at 5 (10) percent level.

Chapter 4

Conclusion

In this thesis, we discuss the estimation techniques for two econometric models, random effects panel data model and simultaneous equations model. We propose a P-splines estimator for random effects panel data model with its asymptotic properties established. Monte Carlo simulation shows that the P-splines estimator consistently performs well in terms of both statistical efficiency and computational efficiency.

For a simultaneous equations model with any number of continuous and binary dependent variables, we derive the corrected asymptotic variance-covariance matrices of two stage estimators that will be of great help for practitioners in econometrics. We also apply our results to empirical data and summarize our findings. It turns out that the difference of the results with corrected and uncorrected errors may be large enough to draw different conclusions.