Estimation of Semi-parametric Functional-coefficient Panel Data Models with Fixed Effects

by

Shaymal C. Halder

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 8, 2020

Keywords: Semi-parametric, smooth coefficient, fixed effect, local linear, LSDV, first difference, time effect

Approved by

Philippe R. Gaillard, Chair, Associate Professor of Statistics
Emir Malikov, Co-chair, Assistant Professor of Economics
David Mark Carpenter, Professor of Statistics
Bertram Zinner, Associate Professor of Mathematics
Kumer Das, Professor of Statistics

Abstract

We consider the problem of estimating a semiparametric varying coefficient panel data model where the unobserved individual effects are correlated with explanatory variables in an unknown arbitrary way using a local linear regression approach. We present a new technique to estimate this model whereby, we locally approximate the fixed-effects-free transformed equation around two different points. Using Monte Carlo simulations, we study potential gains in the finite sample performance and/or the computational time of the proposed estimation procedure over available alternatives under different scenarios. We also consider a conceptually different approach to controlling for unobserved fixed effects in which the fixed effect is modelled as an unknown function of an unordered factor variable indexing individuals. The existing semiparametric estimators for varying-coefficient fixed-effects models exclusively assume one-way fixed effects, typically in the dimension of cross-sectional units. However, more often than not applied researchers wish to control for both the individual and time fixed effects in their panel regressions, with the latter included to account for common unobservable factors correlated with regressors. While rather trivial in a linear model, controlling for time effects by explicitly including time-period dummies as additional regressors does not provide a straight-forward estimation procedure in the case of a semiparametric model. We provide an alternative by extending the Sun et al. (2009) smoothed least-squares dummy variable (LSDV) estimator to the case of a functional-coefficient model with *two-way* fixed effects whereby we allow for unobservable heterogeneity in both dimensions of the data: cross-section and time. Both fixed effects are concentrated out of the model via

locally smoothed two-dimensional within transformation. Simulations show that the estimator performs well in finite samples. We showcase its practical usefulness in two different scenarios.[1]

[1]This chapter includes excerpts from " Halder, S., and Malikov, E. Smoothed LSDV Estimation of Functional-Coefficient Panel Data Models with Two-Way Fixed Effects, Economics Letters. 2020, Volume 192, Article 109239."

Acknowledgments

At first, I would like to take the opportunity to express my sincere gratitude to Dr. Emir Malikov for his advice, guidance, and tremendous supervision from the very early stage of this research as well as giving me extraordinary experiences throughout the work. He always provided me unflinching encouragement and support in various ways to accomplish this project. He always provided valuable insight and direction to the end of this project. I am thankful to Dr. Philippe Gaillard for his throughout support and advice during my Ph.D program in the Department. I am also very much thankful to the remaining dissertation committee members- Drs. Mark Carpenter, Bertram Zinner, and Kumer Das and the University Reader Dr. Aleksandr Vinel for their indispensable advice in this project. This work would not have been possible without the guidance and help of several individuals including Asim Dey and Prasenjit Ghosh. It is a pleasure to convey my gratitude to all of them my humble acknowledgment.

I am thankful to the secretaries and staffs of the Department of Mathematics & Statistics of Auburn University for helping the Department run smoothly and for assisting me in different ways towards the accomplishment of my Ph.D. degree.

Collective and individual acknowledgments are also owed to my graduate friends at Auburn University, whose presence is somehow perpetually refreshing, helpful, memorable, and encouraging during my study at Auburn University.

I would like to express my love and gratitude to my family members, specially my late father, eldest sister Anjana Rani Naha, my wife Lovely Dey, and uncle Sachindra Halder and friends. My parents deserve special mention for their unconditional love, support, and prayers. Finally, I would like to dedicate this work to my mother.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

In statistics and econometrics, a panel data (or longitudinal data) is a multi-dimensional data involving measurements over time. A panel data contains observations of multiple phenomena obtained over multiple time periods for the same firms or individuals.

A linear fixed-effects model is a workhorse of applied research in economics. For example, [19],[17], [18], and [13] are some excellent source of parametric panel data model analysis. The model's broad popularity chiefly stems from its ability (given the availability of panel data) to control for unobservable confounders that may correlate with regressors. However, just like all parametric models, a linear fixed-effects regression is prone to misspecification owing to its reliance on the parametric form assumption (here, linearity) the violation of which may lead to inconsistency and thus misleading inference ; while partially linear semiparametric models may be too restrictive as they only allow for some additive nonlinearities. The semiparametric functional-coefficient[1] fixed-effects model of [1] provides a means to robustify the conventional linear fixed-effects regression by letting its coefficients be unspecified nonparametric functions of relevant contextual variables which, among other benefits, accommodates potential heterogeneity in marginal effects of linear regressors. While not as flexible as a fully nonparametric fixed-effects model [15], such a semiparametric specification is attractive because of its ability to alleviate the so-called "curse of dimensionality" associated with nonparametric estimation and thus to achieve better convergence rates. For instance, [11] considered fixed effects varying coefficient models. [14] , [20] and [36] also

[1]Also sometimes referred to as a "varying-coefficient" or "smooth-coefficient" model.

considered fixed-effects panel data models as well as partially linear fixed-effects panel data models.

We consider the problem of estimating the following varying-coefficient panel data model with individual fixed effects using a local linear regression approach:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}(\mathbf{z}_{it}) + \mu_i + u_{it} \tag{1.1}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$. The covariate vector $\mathbf{z}_{it} = (z_{it,1}, \ldots, z_{it,q})'$ is of dimension q, $\mathbf{x}_{it} = (x_{it,1}, \ldots, x_{it,p})'$ is of dimension p and excludes time-invariant regressors for identification purposes, $\boldsymbol{\beta}(\cdot) = (\beta(\cdot)_1, \ldots, \beta(\cdot)_p)'$ contains p unknown functions, and all other variables are scalars. None of the variables in $\mathbf{x}_{it}$ can be obtained from $\mathbf{z}_{it}$ and vice versa. The random errors u_{it} are assumed to be i.i.d. over i and t with a zero mean and finite variance, and $E[u_{it}|\mu_i, \mathbf{x}_i, \mathbf{z}_i] = 0$, where $\mathbf{x}_i = (\mathbf{x}'_{i1}, \ldots, \mathbf{x}'_{iT})'$ and $\mathbf{z}_i = (\mathbf{x}'_{i1}, \ldots, \mathbf{z}'_{iT})'$. Individual effects μ_i correlate with $\mathbf{z}_{it}$ and/or $\mathbf{x}_{it}$ with an unknown correlation structure. Hence,the model in eq. (1.1) is a fixed-effects model.

When the individual effects μ_i are not correlated with $\mathbf{z}_{it}$ and/or $\mathbf{x}_{it}$ then the the model in eq. (1.1) is a random-effects model.

In chapter 2, we compare the relative finite sample performance (on the basis of root mean squared error, mean absolute error, and computational time) of the two existing estimators and a proposed (new) alternative estimation procedure.

In chapter 3, we consider a conceptually different approach to controlling for unobserved fixed effects in which μ_i is modelled as an unknown function of an unordered factor variable indexing individuals, i.e., $\mu_i = \mu(D_i)$, where D_i is a discrete scalar variable. As its core, such an approach is similar in the spirit to the Least Squares Dummy Variable Approach, whereby fixed effects are modeled via n dummies included as additional regressors, except that we circumvent the need for numerous indicator variables by relying on the ability of kernel estimators to tackle unordered discrete data. Effectively, we can use a single factor variable to deliver the same information. The estimation procedure is two-step a la [37].

In chapter 4, we extend model in eq. (1.1) to also allow for time fixed effects λ_t. We then generalize the Smoothed Least Square Dummy Variable estimator of [1] to accommodate such effects as well and study its finite-sample performance. In this case, we do not however need to assume additivity of λ_t and μ_i.

In chapter 5, we showcase the practical usefulness of our proposed model in chapter 4 in two different scenarios using balanced data as well as unbalanced data. Finally, the chapter 6 is for concluding remarks. [2]

[2]This chapter includes excerpts from " Halder, S., and Malikov, E. Smoothed LSDV Estimation of Functional-Coefficient Panel Data Models with Two-Way Fixed Effects, Economics Letters. 2020, Volume 192, Article 109239."

Chapter 2

Estimation Methodologies

In this chapter, we compare the relative finite sample performance (on the basis of root mean squared error, mean absolute error, and computational time) of the two existing estimators by [1] and [24][23]. Specifically,[1] use the Smoothed Least Square Dummy Variable one-stage approach that locally approximates the original estimating equation around the point $\mathbf{z}_{it} = z$ and then concentrates fixed effects out. [24][23] use a two-stage approach whereby they first remove fixed effects via first differencing and then approximate the transformed equation, now containing both $\boldsymbol{\beta}(\mathbf{z}_{it})$ and $\boldsymbol{\beta}(\mathbf{z}_{it-1})$, around the same two points $(\mathbf{z}_{it}, \mathbf{z}_{it-1}) = (z_1, z_1)$. We also propose a (new) alternative estimation procedure based on a modification of [11] alternative two-stage first-difference approach which, unlike [24][23], approximates the transformed equation around two different points $(\mathbf{z}_{it}, \mathbf{z}_{it-1}) = (z_1, z_2)$ thereby having the potential to significantly reduce the computational time. We then compare finite sample performance (on the basis of root mean squared error, mean absolute error, and computational time) of the proposed estimator with the existing estimators. Using local-linear fitting, the simulation study shows that our proposed estimator is computationally more efficient in finite samples.

2.1 Sun et al (2009) estimator

[1]removes the unknown fixed effects by partialing them out motivated by a least squares dummy variable (LSDV) model in parametric panel data analysis. Writing the model in (1.1) in the matrix from, we have

$$\mathbf{y} = \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(\mathbf{Z})\} + \mathbf{D}\boldsymbol{\mu} + \mathbf{u}, \tag{2.1}$$

where $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_n')'$ and $\mathbf{u} = (\mathbf{u}_1', \ldots, \mathbf{u}_n')'$ are $(nT) \times 1$ vectors, with $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ and $\mathbf{u}_i = (u_{i1}, \ldots, u_{iT})'$. The operator $\mathbb{M}\{\cdot\}$ stacks $\mathbf{x}_{it}'\boldsymbol{\beta}(\mathbf{z}_{it})$ into a $nT \times 1$ vector with the (i, t) subscripts matching those of $\mathbf{y}$ and $\mathbf{u}$. Further, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ is an $n \times 1$ vector of individual fixed effects, and $\mathbf{D} = \mathbf{I}_n \otimes \mathbf{i}_T$ is an $(nT) \times n$ design matrix, where $\mathbf{I}_n$ is an identity matrix of dimension n and $\mathbf{i}_T$ denotes an $T \times 1$ vector of ones.

First, define a $nT \times nT$ diagonal matrix of local kernel weights $\mathbf{W}_H(z) = \text{diag}\{\mathbf{K}_H(\mathbf{z}_1, z), \cdots, \mathbf{K}_H(\mathbf{z}_n, z)$ with $\mathbf{K}_H(\mathbf{z}_i, z) = \text{diag}\{k_H(z_{i1}, z), \cdots, k_H(z_{iT}, z)\}$ being a $T \times T$ diagonal matrix for each i, where $k_H(\mathbf{z}_{it}, z) = k\{\mathbf{H}^{-1}(\mathbf{z}_{it} - z)\}$ is the q-variate product kernel such that $K(u, v) = K(u)K(v)$, where for each u, v, $\int K(u)du = 1$ and $K_h(u) = (1/h)K(u/h)$ and $\mathbf{H} = \text{diag}\{h_1, \ldots, h_q\}$ is a diagonal bandwidth matrix of dimension q. To derive the estimator for unknown functional coefficients, we then solve the following locally weighted least-squares problem:

$$\min_{\boldsymbol{\beta}(z), \boldsymbol{\mu}} [\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\} - \mathbf{D}\boldsymbol{\mu}]'\mathbf{W}_H(z)[\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\} - \mathbf{D}\boldsymbol{\mu}]. \tag{2.2}$$

The first-order condition of (2.2) with respect to individual fixed effects, $\boldsymbol{\mu}$ is

$$\mathbf{D}'\mathbf{W}_H(z)[\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\} - \mathbf{D}\boldsymbol{\mu}] = 0 \tag{2.3}$$

which can be solved for $\widehat{\boldsymbol{\mu}}$, i.e.,

$$\widehat{\boldsymbol{\mu}}(z) =(\mathbf{D}'\mathbf{W}_H(z)\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}_H(z)(\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\}), \tag{2.4}$$

Define the *local* within transformation matrix: $\mathbf{M}_H(z) = \mathbf{I}_{nT} - \mathbf{D}[\mathbf{D}'\mathbf{W}_H(z)\mathbf{D}]^{-1}\mathbf{D}'\mathbf{W}_H(z))$. Then, substituting $\widehat{\boldsymbol{\mu}}(z)$ from (2.4) for $\boldsymbol{\mu}$ in the objective function in (2.2) yields a concentrated locally weighted least-squares problem from which the individual fixed effects are removed:

$$\min_{\boldsymbol{\beta}(z)} (\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\})'\boldsymbol{S}_H(z) (\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\}), \tag{2.5}$$

where $\boldsymbol{S}_H(z) \equiv \mathbf{M}_H(z)'\mathbf{W}_H(z)\mathbf{M}_H(z)$. The riddance of fixed effects from the model is ensured by $\mathbf{M}_H(z)\mathbf{D}\boldsymbol{\mu} = \mathbf{0}_{nT\times 1}$ for all z. Upon a close examination of the local weighting matrix $\boldsymbol{S}_H(z)$, it is evident that $\boldsymbol{\mu}$ is removed via local z-specific kernel-weighted within transformation whereby a kernel-weighted time average is subtracted from each variable.

To operationalize the estimator of $\boldsymbol{\beta}(z)$ from the profiled problem in (2.5), we rely on local-polynomial kernel approximators. For each $s = 1, \ldots, p$, the local Taylor expansion of the unknown functional coefficient around $\mathbf{z}_{it} = z$ is

$$\beta_s(\mathbf{z}_{it}) = \beta_s(z) + (\mathbf{z}_{it} - z)'\nabla_{\mathbf{z}}\beta_s(z) + (\mathbf{z}_{it} - z)'\nabla^2_{\mathbf{z}}\beta_s(z)(\mathbf{z}_{it} - z) + \ldots, \tag{2.6}$$

where $\nabla_{\mathbf{z}}\beta_s(z) = (\partial\beta_s(z)/\partial z_{1,it}, \ldots, \partial\beta_s(z)/\partial z_{q,it})'$ is a $q \times 1$ vector of first-order gradients and $\nabla^2_{\mathbf{z}}\beta_s(z) = \nabla_{\mathbf{z}'}(\nabla_{\mathbf{z}}\beta_s(z))$ is the $q \times q$ Hessian matrix of the second-order derivatives, etc. The first-order (local linear) approximation is arguably the most popular among practitioners [16], and we adopt it here too and so do [1]. Thus, in what follows, we make use of $\beta_s(\mathbf{z}_{it}) \approx \beta_s(z) + (\mathbf{z}_{it} - z)\nabla_{\mathbf{z}}\beta_s(z)$ around $\mathbf{z}_{it} = z$.

Define a $(q+1) \times 1$ vector $\boldsymbol{\theta}_s(z) = (\beta_s(z), \nabla_{\mathbf{z}}\boldsymbol{\beta}_s(z)')'$ of unknown local parameters for each $s = 1, \ldots, p$. Then, the unknown $p \times (q+1)$ parameter matrix is defined as $\boldsymbol{\Theta}(z) = [\boldsymbol{\theta}_1(z) \ \ldots \ \boldsymbol{\theta}_p(z)]'$:

$$\boldsymbol{\Theta}(z) \equiv \begin{bmatrix} \boldsymbol{\theta}_1(z)' \\ \vdots \\ \boldsymbol{\theta}_p(z)' \end{bmatrix} = \begin{bmatrix} \beta_1(z) & \nabla\boldsymbol{\beta}_1(z)' \\ \vdots & \vdots \\ \beta_p(z) & \nabla\boldsymbol{\beta}_p(z)' \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}(z) & \nabla\boldsymbol{\beta}(z)' \end{bmatrix},$$

where the first column of the above matrix is $\boldsymbol{\beta}(\cdot)$ evaluated at z which is of primary interest. Next, define a $(q+1) \times 1$ vector of deviations from z, i.e., $\mathcal{Z}_{it}(z) = (1, (\mathbf{z}_{it} - z)')'$. For $\mathbf{z}_{it}$ close to z, we replace $\boldsymbol{\beta}(\mathbf{z}_{it})$ in (2.5) with $\boldsymbol{\Theta}(z)\mathcal{Z}_{it}(z)$ and obtain the local-linear estimator of functional coefficients $\boldsymbol{\beta}(z)$ from the locally concentrated minimization problem:

$$\min_{\boldsymbol{\Theta}(z)} \left(\mathbf{y} - \mathcal{X}(z)\text{vec}\{\boldsymbol{\Theta}(z)\}\right)' \boldsymbol{S}_H(z) \left(\mathbf{y} - \mathcal{X}(z)\text{vec}\{\boldsymbol{\Theta}(z)\}\right), \tag{2.7}$$

where we stack (by columns) the unknown parameter matrix $\boldsymbol{\Theta}(z)$ into a $p(q+1)\times 1$ vector denoted by the operator vec$\{\cdot\}$, and $\mathcal{X}(z) = (\mathcal{X}_i'(z), \ldots, \mathcal{X}_n'(z))'$ is an $nT \times p(q+1)$ data matrix, with each $T \times p(q+1)$ block given by

$$\mathcal{X}_i(z) = \begin{bmatrix} \mathcal{Z}_{i1}'(z) \otimes \mathbf{x}_{i1}' \\ \vdots \\ \mathcal{Z}_{iT}'(z) \otimes \mathbf{x}_{iT}' \end{bmatrix}.$$

Lastly, solving the first-order condition of (2.7) for the unknown $\boldsymbol{\Theta}(z)$ yields the following local-linear two-way fixed-effects estimator:

$$\text{vec}\left\{\widehat{\boldsymbol{\Theta}}(z)\right\} = \left(\mathcal{X}(z)'\boldsymbol{S}_H(z)\mathcal{X}(z)\right)^{-1} \mathcal{X}(z)'\boldsymbol{S}_H(z)\mathbf{y}. \tag{2.8}$$

2.2 Rodrigue-Poo & Soberon (2014) Estimator

[24] remove the unknown fixed effects, u_i of the model in (1.1) by transforming the model in order to obtain a consistent estimator of the parameters of interest. Any estimation technique suffers from the so called incidental parameters problem, e.g., [34]. A standard solution to this problem is to remove μ_i from (1.1) by taking a transformation, and then estimating the unknown curve through the use of a non-parametric smoother [24]. The simplest approach to remove the unknown fixed effects is to take first differences, i.e.,

$$\Delta y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}(\mathbf{z}_{it}) - \mathbf{x}_{i(t-1)}'\boldsymbol{\beta}(\mathbf{z}_{i(t-1)}) + \Delta u_{it} \tag{2.9}$$

$i = 1, \cdots, n;\ t = 2, \cdots, T$. $\Delta y_{it} = y_{it} - y_{i(t-1)}$ and $\Delta u_{it} = u_{it} - u_{i(t-1)}$. Consider the univariate case, $p = q = 1$. Apply first order Taylor expansion for $(\mathbf{z}_{it})$ and $(\mathbf{z}_{i(t-1)})$ at an interior point $\mathbf{z} = [z_1, z_2]$ such that $\|\mathbf{z}_{it} - z_1\| = o(1)$ and $\|\mathbf{z}_{i(t-1)} - z_2\| = o(1)$ with $z_1 = z_2$

$$\Delta y_{it} \approx \quad \boldsymbol{\beta}(z_1)\mathbf{x_{it}} - \boldsymbol{\beta}(z_2)\mathbf{x_{i(t-1)}} + \boldsymbol{\beta}'(z_1)\mathbf{x_{it}}(\mathbf{z_{it}} - z_1) - \boldsymbol{\beta}'(z_2)\mathbf{x_{i(t-1)}}(\mathbf{z_{i(t-1)}} - z_2)) \quad (2.10)$$

for a given (i, t). So we estimate $\boldsymbol{\beta(z)}$, $\boldsymbol{\beta'(z)}$ by regressing Δy_{it} on the terms $(\mathbf{x}_{it}(\mathbf{z}_{it} - z_1) - \mathbf{x}_{i(t-1)}(\mathbf{z}_{i(t-1)} - z_2))$ with kernel weights. To derive the estimator for unknown functional coefficients, we solve the following locally weighted least-squares problem:

$$\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta y_{it} - \boldsymbol{\beta}(z_1)\mathbf{x_{it}} + \boldsymbol{\beta}(z_2)\mathbf{x_{i(t-1)}} - \boldsymbol{\beta}'(z_1)\mathbf{x_{it}}(\mathbf{z_{it}} - z_1) + \boldsymbol{\beta}'(z_2)\mathbf{x_{i(t-1)}}(\mathbf{z_{i(t-1)}} - z_2))^2 \times$$
$$K((\mathbf{z}_{it} - z_1)/h)K((\mathbf{z}_{i(t-1)} - z_2)/h) \quad (2.11)$$

h is the bandwidth. Here, K **is** $2q$**-variate kernel**. **In contrast, in the case of Sun et al. estimation, the kernel is** q**-variate**

[24] propose a transformation that is a one-step backfitting algorithm in order to achieve the desirable rate of $\sqrt{nTh}$. Denote $\Delta\mathbf{y}_{it}^{(1)}$ the following expression:

$$\Delta y_{it}^{(1)} = \Delta y_{it} + \boldsymbol{\beta}(\mathbf{z}_{i(t-1)})\mathbf{x}_{i(t-1)} \quad (2.12)$$

$i = 1, \cdots, n;\ t = 2, \cdots, T$. Substitute eq. (2.9) into eq. (2.12), we get

$$\Delta y_{it}^{(1)} = \boldsymbol{\beta}(\mathbf{z}_{it})\mathbf{x}_{it} + \Delta u_{it} \quad (2.13)$$

$i = 1, \cdots, n;\ t = 2, \cdots, T$

The eq. (2.13) clearly indicates the estimation of $\boldsymbol{\beta}(.)$ is now a q dimensional problem, and we use local linear least-squares estimation procedure with q-variate kernel weights.

To smooth over fewer variables (q instead of $2q$ as in the first stage), we can achieve a better convergence rate. In eq. (2.12), $\boldsymbol{\beta}\mathbf{x}_{i(t-1)}$ is unknown. We replace it by the initial local linear regression estimator, $\Delta\tilde{y}_{it}^{(1)} = \Delta y_{it} + \hat{\boldsymbol{\beta}}_{\boldsymbol{h}}(\mathbf{z}_{i(t-1)})\mathbf{x}_{i(t-1)}$ with the following regression model:

$$\Delta\tilde{y}_{it}^{(1)} = \boldsymbol{\beta}(\mathbf{z}_{it})\mathbf{x}_{it} + u_{it} \tag{2.14}$$

$i = 1, \cdots, n$; $t = 2, \cdots, T$. Here, $u_{it} = (\hat{\boldsymbol{\beta}}_{\boldsymbol{h}}(\mathbf{z}_{i(t-1)}) - \hat{\boldsymbol{\beta}}(\mathbf{z}_{i(t-1)}))\mathbf{x}_{i(t-1)} + \Delta u_{it}$.

So we estimate $\tilde{\boldsymbol{\beta}}(\cdot)$ with the following weighted local linear regression:

$$\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta\tilde{y}_{it}^{(1)} - \tilde{\boldsymbol{\beta}}(\mathbf{z})\mathbf{x}_{it} - \tilde{\boldsymbol{\beta}'}(\mathbf{z})\mathbf{x}_{it}(\mathbf{z}_{it} - \mathbf{z}))^2 K_{\bar{h}}(\mathbf{z}_{it} - \mathbf{z}) \tag{2.15}$$

$\tilde{\boldsymbol{\beta}}(\mathbf{z})$ and $\tilde{\boldsymbol{\beta}'}(\mathbf{z})$ are the estimators of $\boldsymbol{\beta}(\mathbf{z})$ and $\boldsymbol{\beta}'(\mathbf{z})$ respectively. The direct application of local linear regression techniques to first-differencing transformations in panel data models renders biased estimators and the bias does not degenerate, even with large samples [44]. Using a higher dimensional kernel weight, [24] the estimation technique overcomes the problem of non-vanishing bias. However, as expected, the variance term becomes larger.

2.3 Modified Rodrigue-Poo & Soberon Estimator

We can improve the computational time of the [24] first difference approach. We propose a modified estimator based on [11] alternative two-stage first-difference approach which, unlike [24], approximates the transformed equation around two different points $(\mathbf{z}_{it}, \mathbf{z}_{it-1}) = (z_1, z_2)$, $z_1 \neq z_2$.

First take the first differences of our model of eq. (1.1) to remove the heterogeneity of unknown form:

$$\Delta y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}(\mathbf{z}_{it}) - \mathbf{x}'_{i(t-1)}\boldsymbol{\beta}(\mathbf{z}_{i(t-1)}) + \Delta u_{it} \tag{2.16}$$

$i = 1, \cdots, n$; $t = 2, \cdots, T$. $\Delta y_{it} = y_{it} - y_{i(t-1)}$ and $\Delta u_{it} = u_{it} - u_{i(t-1)}$.

Consider the univariate case, $p = q = 1$. Apply Taylor expansion for $(\mathbf{z}_{it})$ and $(\mathbf{z}_{i(t-1)})$ at an interior point $\mathbf{z} = [z_1, z_2]$ with $z_1 \neq z_2$

$$\Delta y_{it} \approx \quad \boldsymbol{\beta}(z_1)\mathbf{x_{it}} - \boldsymbol{\beta}(z_2)\mathbf{x_{i(t-1)}} + \boldsymbol{\beta}'(z_1)\mathbf{x_{it}}(\mathbf{z_{it}} - z_1) - \boldsymbol{\beta}'(z_2)\mathbf{x_{i(t-1)}}(\mathbf{z_{i(t-1)}} - z_2)) \tag{2.17}$$

for a given (i, t). So we estimate $\boldsymbol{\beta}(\mathbf{z})$, $\boldsymbol{\beta}'(\mathbf{z})$ by regressing Δy_{it} on the terms $(\mathbf{x}_{it}(\mathbf{z}_{it} - z_1) - \mathbf{x}_{i(t-1)}(\mathbf{z}_{i(t-1)} - z_2))$ with kernel weights. To derive the estimator for unknown functional coefficients, we solve the following locally weighted least-squares problem:

$$\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta y_{it} - \boldsymbol{\beta}(z_1)\mathbf{x_{it}} + \boldsymbol{\beta}(z_2)\mathbf{x_{i(t-1)}} - \boldsymbol{\beta}'(z_1)\mathbf{x_{it}}(\mathbf{z_{it}} - z_1) + \boldsymbol{\beta}'(z_2)\mathbf{x_{i(t-1)}}(\mathbf{z_{i(t-1)}} - z_2))^2 \times \\ K((\mathbf{z}_{it} - z_1)/h)K((\mathbf{z}_{i(t-1)} - z_2)/h) \tag{2.18}$$

h is the bandwidth.

So comparing the eq. (2.11) and eq. (2.18), we see the contrast. The original [24] estimator in eq. (2.11) requires nT repeated approximations/estimations, whereas the proposed modified estimator in eq. (2.18)) asks only for $n(T-1)$ such approximations/estimations. So we can intuitively explain that the proposed modification is more computationally faster.

The second stage estimator of our proposed Modified Rodriguez-Poo & Soberon estimator follows the same steps as did by [24]. Denote $\Delta y_{it}^{(1)}$ the following expression:

$$\Delta y_{it}^{(1)} = \Delta y_{it} + \boldsymbol{\beta}(\mathbf{z}_{i(t-1)})\mathbf{x}_{i(t-1)} \tag{2.19}$$

$i = 1, \cdots, n$; $t = 2, \cdots, T$. Substitute eq. (2.16) into eq. (2.19), we get

$$\Delta y_{it}^{(1)} = \boldsymbol{\beta}(\mathbf{z}_{it})\mathbf{x}_{it} + \Delta u_{it} \tag{2.20}$$

$i = 1, \cdots, n$; $t = 2, \cdots, T$

The eq. (2.20) indicates the estimation of $\boldsymbol{\beta}(.)$ becomes a q dimensional problem, and we use local linear least-squares estimation procedure with q-variate kernel weights.

To smooth over fewer variables (q instead of $2q$ as in the first stage), we can achieve a better convergence rate. In eq. (2.19), $\boldsymbol{\beta}\mathbf{x}_{i(t-1)}$ is unknown. We replace it by the initial local linear regression estimator, $\Delta \tilde{y}_{it}^{(1)} = \Delta y_{it} + \hat{\boldsymbol{\beta}}_{\boldsymbol{h}}(\mathbf{z}_{i(t-1)})\mathbf{x}_{i(t-1)}$ with the following regression model:

$$\Delta \tilde{y}_{it}^{(1)} = \boldsymbol{\beta}(\mathbf{z}_{it})\mathbf{x}_{it} + u_{it} \tag{2.21}$$

$i = 1, \cdots, n$; $t = 2, \cdots, T$. Here, $u_{it} = (\hat{\boldsymbol{\beta}}_{\boldsymbol{h}}(\mathbf{z}_{i(t-1)}) - \hat{\boldsymbol{\beta}}(\mathbf{z}_{i(t-1)}))\mathbf{x}_{i(t-1)} + \Delta u_{it}$.

So we estimate $\tilde{\boldsymbol{\beta}}(\cdot)$ with the following weighted local linear regression:

$$\sum_{i=1}^{n}\sum_{t=2}^{T}(\Delta \tilde{y}_{it}^{(1)} - \tilde{\boldsymbol{\beta}}(z)\mathbf{x}_{it} - \tilde{\boldsymbol{\beta}}'(z)\mathbf{x}_{it}(\mathbf{z}_{it} - z))^2 K_{\bar{h}}(\mathbf{z}_{it} - z) \tag{2.22}$$

$\tilde{\boldsymbol{\beta}}(\mathbf{z})$ and $\tilde{\boldsymbol{\beta}'}(\mathbf{z})$ are the estimators of $\boldsymbol{\beta}(\mathbf{z})$ and $\boldsymbol{\beta}'(\mathbf{z})$ respectively.

2.4 Simulation Study

We study the finite-sample performance of the model (1.1) in a series of Monte Carlo experiments of the [1], [24], and our proposed Modified Rodriguez-Poo & Soberon estimator.

First, we consider the Data Generating Process (DGP) for univariate case with $p = q = 1$ whereby $y_{it} = x_{it}\beta(z_{it}) + \mu_i + u_{it}$, where the variables are drawn as follows: $z_{it} = 0.5(\omega_{it} + \omega_{it-1})$, where $\omega_{it} \sim$ i.i.d. $\mathcal{U}(0, 0.5\pi)$; $x_{it} = 0.5(bz_{it} + x_{it-1}) + \zeta_{it}$, where $\zeta_{it} \sim$ i.i.d. $\mathcal{N}(0, 1)$; and $u_{it} \sim$ i.i.d. $\mathcal{N}(0, 0.5)$. We consider two cases: (1) x_{it} and z_{it} are correlated with $b = 1$ and (2) x_{it} and z_{it} are uncorrelated with $b = 0$. The outcome is generated with the following specification of individual effects: $\mu_i = c(\overline{z}_i + \overline{x}_i) + \rho_i$ with $\rho_i \sim$ i.i.d. $\mathcal{N}(0, 0.5)$, where c controls the degree of correlation with regressors. Here, we set $c = 0.5$ for "fixed" effects. The functional coefficient is specified as $\beta(z_{it}) = \sin(\pi z_{it})$.

We consider cross-sectional sample sizes $n = \{50,\ 100,\ 200\}$ with the number of time periods $T = \{3,\ 5\}$. For each (n, T), we simulate the model 500 times. We use the popular [9] rule-of-thumb bandwidth for the smoothing variables. The kernel function of choice is second-order Gaussian. For each simulation, we compute the average (over z_{it}) root mean squared error (RMSE) and the average (over z_{it}) mean absolute error (MAE) for each functional coefficient function:

$$RMSE\left(\widehat{\beta}(\cdot)\right) = \sqrt{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\widehat{\beta}(z_{it}) - \beta(z_{it})\right]^2}$$

$$MAE\left(\widehat{\beta}(\cdot)\right) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left|\widehat{\beta}(z_{it}) - \beta(z_{it})\right|,$$

and then report their respective averages computed over 500 simulations in Table 2.1–2.6. We also report the computational time for all estimators.

The results in Table 2.1–2.6 are encouraging and indicate that, in all cases, the estimation of $\beta(\cdot)$ of [1], [24], and our proposed Modified Rodriguez-Poo & Soberon estimator

Table 2.1. Simulation results for the Sun et al. (2009) estimator ($p = 1$, $q = 1$)

	$T = 3$			$T = 5$		
	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
	Case 1: x and z are correlated					
RMSE	0.1721	0.1338	0.1045	0.1179	0.0940	0.0739
MAE	0.1201	0.0915	0.0707	0.0803	0.0624	0.0468
Com.time (s)	804.67	6,085.24	66,714.25	3,257.09	30,182.19	453,627.8
	Case 2: x and z are uncorrelated					
RMSE	0.1822	0.1409	0.1099			
MAE	0.1239	0.0939	0.0727			
Com.time (s)	765.48	5,816.55	62,066.40			

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$.

becomes more stable as the sample size increases for both fixed T and fixed n. Both the RMSE and MAE decline significantly.

Next, we examine the performance of the estimators in mutivariate case. We increase the number of variables that enter the model nonparametrically. The fixed-effects DGP with $p = 1$ regressor but $q = 2$ smoothing variables is $y_{it} = x_{it}\beta(z_{it}) + \mu_i + u_{it}$, where $z_{l,it} = 0.5(\omega_{l,it} + \omega_{l,it-1})$ for $l = 1, \ldots, q$; $x_{it} = 0.5(z_{q,it} + x_{it-1}) + \zeta_{it}$; $\mu_i = c(0.5\overline{z}_{1,i} + 0.5\overline{z}_{2,i} + \overline{x}_i) + \rho_i$ with $c = 0.5$. The remaining random terms $\omega_{l,it}$, ζ_{it}, u_{it}, ρ_i and ϱ_t are drawn as before. The functional coefficient is specified as $\beta(z_{it}) = 1 + z_{1,it}z_{2,it} + z_{2,it}^2$. The corresponding results are reported in Table 2.7–2.12.

2.5 Comparison of Computation Time

We perform simulation study in finite samples and compare the computational time for existing estimators of semiparametric functional-coefficient panel data models with fixed effects by [1], [24], and our proposed Modified Rodriguez-Poo & Soberon estimator. We used a system with the Core i7, 8.00 GB memory with 64-bit Operating System, x64-based processor to compute the estimators. The computational time (in seconds) to evaluate the existing estimators and our proposed estimator is reported in Table 2.13 for univariate case. The percentage time saved relative to the base line estimator (Sun et al. estimator)

Table 2.2. Simulation results for the Rodriguez-Poo & Soberon (2014) Two-Stage estimator ($p = 1$, $q = 1$)

		$T = 3$			$T = 5$		
		$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
Case 1: x and z are correlated							
1st Stage	RMSE	0.2273	0.1815	0.1373	0.1741	0.1430	0.1100
	MAE	0.1540	0.1203	0.0923	0.1183	0.0931	0.0714
	Com.time (s)	544.03	2,204.69	9,044.61	1,450.16	6,318.42	28,334.03
2nd Stage	RMSE	0.5036	0.4612	0.4349	0.4495	0.4297	0.4283
	MAE	0.4052	0.3832	0.3764	0.3743	0.3684	0.3761
	Com.time (s)	943.24	3,991	16,576.89	3,456.47	14,451.92	61,168.94
Case 2: x and z are uncorrelated							
1st Stage	RMSE	0.2411	0.1877	0.1446			
	MAE	0.1585	0.1226	0.0942			
	Com.time (s)	524.48	2,079.22	8,747.26			
2nd Stage	RMSE	0.5337	0.4907	0.4654			
	MAE	0.4269	0.4071	0.4004			
	Com.time (s)	946.51	3,972.37	15,257.72			

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$.

is also reported. The Table 2.14 results for multivariate case. In all cases, our proposed Modified Rodiriguez-Poo & Soberon estimator takes less time to compute the estimator. The first-stage estimator actually performs better than the second step, as can easily be seen from the RMSE results of the [24] estimator and our proposed Modified Rodriguez-Poo & Soberon estimator . So we mainly focus on the performance of computational time gains of the first-stage estimator. In all scenarios, our proposed estimator performs better and the computation gains become more significant when the the sample size n increases. So our proposed estimator is computationally more efficient. [1]

[1] This chapter includes excerpts from " Halder, S., and Malikov, E. Smoothed LSDV Estimation of Functional-Coefficient Panel Data Models with Two-Way Fixed Effects, Economics Letters. 2020, Volume 192, Article 109239."

Table 2.3. Simulation results for the Proposed Modified Rodriguez-Poo & Soberon Two-Stage estimator ($p = 1$, $q = 1$)

		$T = 3$			$T = 5$		
		$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
		Case 1: x and z are correlated					
1st Stage	RMSE	0.2739	0.2094	0.1629	0.1903	0.1500	0.1159
	MAE	0.1841	0.1390	0.1065	0.1269	0.0985	0.0754
	Com.time (s)	412.60	1,721.67	6,997.58	1,839.87	7,110.37	29,297.92
2nd Stage	RMSE	0.5128	0.4642	0.4445	0.4544	0.4321	0.4322
	MAE	0.4164	0.3897	0.3845	0.3821	0.3760	0.3825
	Com.time (s)	857.28	3,430.39	14,056.90	3,446.44	14,411.68	59,916.86
		Case 2: x and z are uncorrelated					
1st Stage	RMSE	0.2918	0.2199	0.1688			
	MAE	0.1937	0.1457	0.1118			
	Com.time (s)	398.94	1,573.57	6,709.93			
2nd Stage	RMSE	0.5473	0.4937	0.4693			
	MAE	0.4384	0.4122	0.4053			
	Com.time (s)	810.52	3,273.68	13,165.84			

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$.

Table 2.4. Simulation results for the Sun et al. (2009) estimator ($p = 1$, $q = 1$)

	$n = 50$	$n = 100$	$n = 200$
	Case 1: x and z are correlated		
RMSE	0.2897	0.2239	0.1745
MAE	0.2035	0.1540	0.1192
Com.time (s)	765.46	5,711.00	62,689.59
	Case 2: x and z are uncorrelated		
RMSE	0.3259	0.2499	0.1953
MAE	0.2203	0.1653	0.1285
Com.time (s)	739.81	5,741.03	62,533.00

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$. $T = 3$ throughout. Here, we change the variability of $\zeta_{it} = 0.5$. $\zeta_{it} \sim$ i.i.d. $\mathcal{N}(0, 0.25)$

Table 2.5. Simulation results for the Rodriguez-Poo & Soberon (2014) Two-Stage estimator ($p = 1$, $q = 1$)

		$n = 50$	$n = 100$	$n = 200$
		Case 1: x and z are correlated		
1st Stage	RMSE	0.3933	0.3151	0.2406
	MAE	0.2727	0.2146	0.1649
	Com.time (s)	484.52	1,991.31	8,424.25
2nd Stage	RMSE	0.5842	0.5029	0.4391
	MAE	0.4422	0.3885	0.3604
	Com.time (s)	902.46	3,698.20	15,298.67
		Case 2: x and z are uncorrelated		
1st Stage	RMSE	0.4388	0.3545	0.2646
	MAE	0.2932	0.2283	0.1749
	Com.time (s)	517.83	2,036.15	8,691.92
2nd Stage	RMSE	0.6784	0.5894	0.5041
	MAE	0.5082	0.4435	0.4094
	Com.time (s)	894.36	3,701.02	15,475.75

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$. $T = 3$ throughout. Here, we change the variability of $\zeta_{it} = 0.5$. $\zeta_{it} \sim$ i.i.d. $\mathcal{N}(0, 0.25)$

Table 2.6. Simulation results for the Proposed Modified Rodriguez-Poo & Soberon Two-Stage estimator ($p = 1$, $q = 1$)

		$n = 50$	$n = 100$	$n = 200$
		Case 1: x and z are correlated		
1st Stage	RMSE	0.4539	0.3497	0.2708
	MAE	0.3144	0.2395	0.1845
	Com.time (s)	385.36	1,588.46	6,570.97
2nd Stage	ARMSE	0.5825	0.4897	0.4385
	AMAD	0.4462	0.3855	0.3638
	Com.time (s)	744.57	3,197.89	13,158.37
		Case 2: x and z are uncorrelated		
1st Stage	RMSE	0.5305	0.3967	0.3073
	MAE	0.3577	0.2683	0.2067
	Com.time (s)	391.54	1,666.42	6,769.00
2nd Stage	RMSE	0.6980	0.5735	0.4999
	MAE	0.5234	0.4405	0.4069
	Com.time (s)	803.96	3,277.07	13,620.36

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$. $T = 3$ throughout. Here, we change the variability of $\zeta_{it} = 0.5$. $\zeta_{it} \sim$ i.i.d. $\mathcal{N}(0, 0.25)$

Table 2.7. Simulation results for the Sun et al. (2009) estimator ($p = 1$, $q = 2$)

	$n = 50$	$n = 100$	$n = 200$
	Case 1: x and z are correlated		
RMSE	0.3292	0.2711	0.2293
MAE	0.2065	0.1647	0.1353
Com.time (s)	763.75	5,696.14	65,100.73
	Case 2: x and z are uncorrelated		
RMSE	0.3574	0.2919	0.2411
MAE	0.2182	0.1725	0.1408
Com.time (s)	766.22	5,827.25	65,944.43

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$. $T = 3$ throughout.

Table 2.8. Simulation results for the Rodriguez-Poo & Soberon (2014) Two-Stage estimator ($p = 1$, $q = 2$)

		$n = 50$	$n = 100$	$n = 200$
		Case 1: x and z are correlated		
	RMSE	0.4709	0.4118	0.3555
1st Stage	MAE	0.2888	0.2435	0.2047
	Com.time (s)	500.66	2,018.57	8,503.10
	RMSE	1.5989	1.3657	1.2453
2nd Stage	MAE	1.1706	1.0637	1.0101
	Com.time (s)	911.95	3,742.94	15,539.04
		Case 2: x and z are uncorrelated		
	RMSE	0.5324	0.4521	0.3769
1st Stage	MAE	0.3071	0.2561	0.2120
	Com.time (s)	517.73	2,172.25	8,973.36
	RMSE	1.8419	1.6526	1.5425
2nd Stage	MAE	1.4184	1.3216	1.2744
	Com.time (s)	945.72	3,796.75	15,822.31

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$. $T = 3$ throughout.

Table 2.9. Simulation results for the Proposed Modified Rodriguez-Poo & Soberon Two-Stage estimator ($p = 1$, $q = 2$)

		$n = 50$	$n = 100$	$n = 200$
		Case 1: x and z are correlated		
1st Stage	RMSE	1.0668	0.9512	0.8585
	MAE	0.6554	0.5718	0.5077
	Com.time (s)	398.98	1,649.39	6,808.45
2nd Stage	RMSE	1.7396	1.4514	1.3108
	MAE	1.2671	1.1171	1.0612
	Com.time (s)	820.46	3,246.17	13,676.62
		Case 2: x and z are uncorrelated		
1st Stage	RMSE	1.2075	1.0889	0.9974
	MAE	0.7292	0.6388	0.5681
	Com.time (s)	428.36	1,705.93	6,997.98
2nd Stage	RMSE	2.0079	1.7511	1.6030
	MAE	1.5109	1.3638	1.3066
	Com.time (s)	858.82	3,474.24	14,394.41

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$. $T = 3$ throughout.

Table 2.10. Simulation results for the Sun et al. (2009) estimator ($p = 1$, $q = 2$)

	$n = 50$	$n = 100$	$n = 200$
	Case 1: x and z are correlated		
RMSE	0.5533	0.4525	0.3837
MAE	0.3508	0.2844	0.2329
Com.time (s)	787.78	5,823.62	64,066.79
	Case 2: x and z are uncorrelated		
RMSE	0.6599	0.5313	0.4416
MAE	0.3999	0.3181	0.2590
Com.time (s)	792.11	5,925.38	66,201.60

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$. $T = 3$ throughout. Here, we change the variability of $\zeta_{it} = 0.5$. $\zeta_{it} \sim$ i.i.d. $\mathcal{N}(0, 0.25)$

Table 2.11. Simulation results for the Rodriguez-Poo & Soberon (2014) Two-Stage estimator ($p = 1,\ q = 2$)

		$n = 50$	$n = 100$	$n = 200$
	Case 1: x and z are correlated			
	RMSE	0.8265	0.7237	0.6064
1st Stage	MAE	0.5077	0.4319	0.3641
	Com.time (s)	528.86	2,189.22	8,933.40
	RMSE	1.4330	1.2175	1.0442
2nd Stage	MAE	0.9723	0.8281	0.7416
	Com.time (s)	938.14	3,837.67	16,025.62
	Case 2: x and z are uncorrelated			
	RMSE	0.9649	0.8437	0.7197
1st Stage	MAE	0.5713	0.4812	0.4005
	Com.time (s)	538.46	2,126.70	8,913.52
	RMSE	1.9643	1.7048	1.5879
2nd Stage	MAE	1.4219	1.2747	1.2035
	Com.time (s)	956.47	3,895.80	16,243.70

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$. $T = 3$ throughout. Here, we change the variability of $\zeta_{it} = 0.5$. $\zeta_{it} \sim$ i.i.d. $\mathcal{N}(0, 0.25)$

Table 2.12. Simulation results for the Proposed Modified Rodriguez-Poo & Soberon Two-Stage estimator ($p = 1,\ q = 2$)

		$n = 50$	$n = 100$	$n = 200$
	Case 1: x and z are correlated			
	RMSE	1.6756	1.4885	1.3349
1st Stage	MAE	1.0202	0.8915	0.7945
	Com.time (s)	432.58	1,732.94	7,184.98
	RMSE	1.6364	1.3565	1.1618
2nd Stage	MAE	1.1092	0.9183	0.8130
	Com.time (s)	875.65	3,427.75	14,228.36
	Case 2: x and z are uncorrelated			
	RMSE	2.3030	2.0795	1.8686
1st Stage	MAE	1.3762	1.2013	1.0622
	Com.time (s)	426.58	1,720.35	7,051.96
	RMSE	2.3712	1.9677	1.7351
2nd Stage	MAE	1.6539	1.3817	1.2590
	Com.time (s)	865.96	3,495.36	14,092.38

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$. $T = 3$ throughout. Here, we change the variability of $\zeta_{it} = 0.5$. $\zeta_{it} \sim$ i.i.d. $\mathcal{N}(0, 0.25)$

Table 2.13. Comparison of Com.Time of existing estimators and our proposed estimators for univariate case

				% Time Saved		
	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
Sun et. al. estimator	804.67	6,085.24	66,714.25			
R-P estimator	544.03	2,204.69	9,044.61			
	943.24	3,991	16,576.89			
Proposed Modified R-P estimator	412.60	1,721.67	6,997.58	48.72	71.71	89.51
	857.28	3,430.39	14,056.90			

Reported are the results for the computational time (s) of the existing estimators and our proposed estimator. T=3 throughout

Table 2.14. Comparison of Com. Time of existing estimators and our proposed estimators for multivariate case

				% Time Saved		
	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
Sun et. al. estimator	763.75	5,696.14	65,100.73			
R-P estimator	500.66	2,018.57	8,503.10			
	911.95	3,742.94	15,539.04			
Proposed Modified R-P estimator	398.98	1,649.39	6,808.45	47.76	71.04	89.54
	820.46	3,246.17	13,676.62			

Reported are the results for the computational time (s) of the existing estimators and our proposed estimator. T=3 throughout

Chapter 3

Semiparametric Methods with Discrete Covariates

In this chapter, we construct nonparametric estimators for a regression model in the presence of unordered discrete variable. In the literarure there have been a handful of studies use discrete regressors in nonparametric regression, e.g. [35] , [42], [43]. However, it is naive to use a discrete variable to control for fixed effects. We perform simulation study and the estimates are not consistent for fixed T as n increases.

3.1 The Model

We consider a conceptually different approach to controlling for unobserved fixed effects in which μ_i is modelled as an unknown function of an unordered factor variable indexing individuals, i.e., $\mu_i = \mu(D_i)$, where D_i is a discrete scalar variable. As its core, such an approach is similar in the spirit to the Least Squares Dummy Variable Approach, whereby fixed effects are modeled via n dummies included as additional regressors, except that we circumvent the need for numerous indicator variables by relying on the ability of kernel estimators to tackle unordered discrete data. Effectively, we can use a single factor variable to deliver the same information. The estimation procedure is two-step a la [37]. Introducing an unordered factor variable indexing individuals, i.e., $\mu_i = \mu(D_i)$, where D_i is a discrete scalar variable in our model in eq. (1.1):

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}(\mathbf{z}_{it}) + \mu(D_i) + u_{it} \tag{3.1}$$

So, μ_i is modelled as an unknown function of an unordered factor variable indexing individuals, i.e., $\mu_i = \mu(D_i)$, where D_i is a discrete scalar variable. [36] developed the unordered kernel functions :

$$l(X_i, x, \lambda) = \begin{cases} 1, & X_i = x \\ \lambda, & X_i \neq x \end{cases}$$

where, $\lambda \in [0, 1]$ is an indicator function.

3.2 Estimation

3.2.1 Single-step Estimator

In single step estimator, we approximate the model in (3.1) around $(z_{it}, D_i) = (z, D)$, where D_i is a discrete scalar variable.

3.2.2 Two-step Estimator

The estimation procedure of eq. (3.1) is two-step a la [37]. First, the model is transformed to remove the unknown function then $\boldsymbol{\beta}$ is estimated using OLS. The way we do this is to take the conditional expectation of eq. (3.1) with respect to z_{it}, D_i.

$$E[y_{it}|\mathbf{z}_{it}, D_i] = E[\mathbf{x}_{it}|\mathbf{z}_{it}, D_i]'\boldsymbol{\beta}(\mathbf{z}_{it}) + \mu(D_i) \tag{3.2}$$

Substitute eq. (3.2) from eq. (3.1):

$$y_{it} - E[y_{it}|\mathbf{z}_{it}, D_i] = (\mathbf{x}_{it} - E[\mathbf{x}_{it}|\mathbf{z}_{it}, D_i])'\boldsymbol{\beta}(\mathbf{z}_{it}) + u_{it} \tag{3.3}$$

1st Step:

If we know $E[y_{it}|\mathbf{z}_{it}, D_i]$ and $E[\mathbf{x}_{it}|\mathbf{z}_{it}, D_i]$, we proceed with OLS to estimate $\boldsymbol{\beta}$. In practice $E[y_{it}|\mathbf{z}_{it}, D_i]$ and $E[\mathbf{x}_{it}|\mathbf{z}_{it}, D_i]$ are unknown and must be estimated. [37] suggests to use Local Constant Least Square (LCLS) Estimator to estimate each conditional mean separately, where h is the Silverman bandwidth.

2nd Step:

$$y_{it} - E[y_{it}|\mathbf{z}_{it}, D_i] = (\mathbf{x}_{it} - E[\mathbf{x}_{it}|\mathbf{z}_{it}, D_i])'\boldsymbol{\beta}(\mathbf{z}_{it}) + u_{it}$$

We estimate eq. (3.3) via Local Linear Least Square (LLLS) Estimator taking an approximation around $z_{it} = z$.

3.3 Simulation Study

We study the finite-sample performance of the model (3.1) in a series of Monte Carlo experiments.

We consider the DGP for univariate case with $p = q = 1$ whereby $y_{it} = x_{it}\beta(z_{it}) + \mu_i + u_{it}$, where the variables are drawn as follows: $z_{it} = 0.5(\omega_{it} + \omega_{it-1})$, where $\omega_{it} \sim$ i.i.d. $\mathcal{U}(0, 0.5\pi)$; $x_{it} = 0.5(z_{it} + x_{it-1}) + \zeta_{it}$, where $\zeta_{it} \sim$ i.i.d. $\mathcal{N}(0, 1)$; and $u_{it} \sim$ i.i.d. $\mathcal{N}(0, 0.5)$. The outcome is generated with the following specification of individual effects: $\mu_i = c(\overline{z}_i + \overline{x}_i) + \rho_i$ with $\rho_i \sim$ i.i.d. $\mathcal{N}(0, 0.5)$, where c controls the degree of correlation with regressors. We estimate fixed effects as well as random effects. Here, we set $c = 0.5$ for "fixed" effects and $c = 0$ for "random" effects. The functional coefficient is specified as $\beta(z_{it}) = \sin(\pi z_{it})$.

We consider cross-sectional sample sizes $n = \{50,\ 100,\ 200\}$ with the number of time periods $T = \{3, 5,\ 7\}$. For each (n, T), we simulate the model 500 times. We use the popular [9] rule-of-thumb bandwidth for the smoothing variables. The kernel function is [2]'s kernel for discrete unordered random variable. We consider the indicator function, $\lambda = 0.5$. For each simulation, we compute the bias, average (over z_{it}) root mean squared error (RMSE), and the average (over z_{it}) mean absolute error (MAE) for each functional coefficient function and then report their respective averages computed over 500 simulations in Table 3.1–3.4. We report results for both fixed effects as well as random effects estimator.

Table 3.1. Simulation study of Fixed effect one-step Discrete Variable estimator

	$T=3$			$T=5$			$T=7$		
	n=50	n=100	n=200	n=50	n=100	n=200	n=50	n=100	n=200
BIAS	-0.2792	-0.2832	-0.2903	-0.2030	-0.2105	-0.2109	-0.1573	-0.1605	-0.1644
RMSE	0.3123	0.3004	0.2998	0.2300	0.2259	0.2193	0.1817	0.1756	0.1721
MAE	0.2858	0.2860	0.2916	0.2089	0.2131	0.2123	0.1630	0.1637	0.1656

Reported are the results for the functional coefficient estimator.

3.4 Concluding Remarks

The simulation results in Table 3.1–3.4 indicate that the estimation $\beta(\cdot)$ are consistent for fixed n as T increases. The estimates are also consistent when both n as T increases. However, the estimates are not consistent for fixed T as n increases. This is due to incidental parameters problem in nonlinear model. Fixed effects generally inconsistent in nonlinear model as n grows with T fixed. In a linear model, least squares treating the additive constant, μ_i as a parameter to be estimated is consistent. Incidental parameters problem in nonlinear model is caused by only having T observations to estimate each μ_i , so that as n grows the estimate of μ_i remains random. In linear models this randomness gets "averaged out." In nonlinear models it does not. This highlights the caution the practitioners ought to use a discrete variable to control for fixed effects. [1]

[1] This chapter includes excerpts from " Halder, S., and Malikov, E. Smoothed LSDV Estimation of Functional-Coefficient Panel Data Models with Two-Way Fixed Effects, Economics Letters. 2020, Volume 192, Article 109239."

Table 3.2. Simulation study of Random Effect one-step Discrete Variable estimator

	$T=3$			$T=5$			$T=7$		
	n=50	n=100	n=200	n=50	n=100	n=200	n=50	n=100	n=200
BIAS	0.0189	0.0189	0.0124	0.0134	0.0103	0.0108	0.0161	0.0128	0.0093
RMSE	0.1326	0.0975	0.0721	0.1028	0.0778	0.0580	0.0877	0.0695	0.0504
MAE	0.0951	0.0709	0.0518	0.0739	0.0568	0.0418	0.0631	0.0498	0.0362

Reported are the results for the functional coefficient estimator.

Table 3.3. Simulation study of Fixed Effect two-step Discrete Variable estimator

	$T=3$			$T=5$			$T=7$		
	n=50	n=100	n=200	n=50	n=100	n=200	n=50	n=100	n=200
BIAS	-0.2667	-0.2737	-0.2822	-0.1916	-0.2017	-0.2037	-0.1473	-0.1528	-0.1583
RMSE	0.2987	0.2906	0.2917	0.2199	0.2175	0.2124	0.1727	0.1679	0.1666
MAE	0.2734	0.2766	0.2835	0.1982	0.2044	0.2052	0.1540	0.1562	0.1596

Reported are the results for the functional coefficient estimator.

Table 3.4. Simulation study of Random Effect two-step Discrete Variable estimator

	$T=3$			$T=5$			$T=7$		
	n=50	n=100	n=200	n=50	n=100	n=200	n=50	n=100	n=200
BIAS	0.0179	0.0184	0.0127	0.0139	0.0106	0.0108	0.0163	0.0132	0.0094
RMSE	0.1317	0.0978	0.0729	0.1050	0.0791	0.0592	0.0894	0.0702	0.0512
MAE	0.0952	0.0709	0.0523	0.0749	0.0572	0.0422	0.0635	0.0501	0.0364

Reported are the results for the functional coefficient estimator.

Chapter 4

Two-way Fixed Effect Estimator

4.1 Introduction

The existing semiparametric estimators for varying-coefficient fixed-effects models [1, 24, 25], however, exclusively assume *one*-way fixed effects, typically in the dimension of individual cross-sectional units.[1] This is unfortunate because more often than not researchers wish to control for *both* unit- and time-specific fixed effects in their panel regressions, especially with the widespread popularity of the difference-in-difference based identification strategies. In addition to the conventional time-invariant individual effects, the time effects are included with the intent to control for common unobservable factors correlated with regressors. In linear models, this is rather simple. Since most microeconomic studies use short panels ($n \ll T$), practitioners customarily control for time effects by explicitly including time-period dummies as additional regressors, whereas the individual fixed effects are usually transformed out of the equation by either the within or first-difference transformation. The same procedure is however not as trivial in the case of a semiparametric functional-coefficient model because the direct estimation of constant coefficients on such time-period dummies renders the model a partially linear, functional-coefficient regression thereby necessitating more than a single-step estimation which, besides theoretical complications, is also more computationally demanding.

To fill this practically important gap, we contribute to the literature by extending the [1] smoothed least-squares dummy variable (SLSDV) estimator to the case of a functional-coefficient panel data model with *two*-way fixed effects whereby we allow for unobservable

[1] The same applies to other types of semiparametric fixed-effects panel data models such as a partially linear specification, see [29, 20, 10].

heterogeneity in both dimensions of the data: cross-section and time. Both fixed effects are (asymptotically) concentrated out of the model via *local* kernel-smoothed two-dimensional within transformation. One of the benefits of the proposed estimation procedure is that the unknown functional coefficients are estimated in a single step. Using local-linear fitting, the simulation experiments show that the two-way SLSDV estimator performs well in finite samples. We also showcase its practical usefulness in two different scenarios.

4.2 The Model

We consider the problem of estimating the following semiparametric functional-coefficient panel data model with *two*-way fixed effects:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}(\mathbf{z}_{it}) + \mu_i + \lambda_t + u_{it}, \tag{4.1}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$. The outcome y_{it} is a scalar; the covariate column vector $\mathbf{x}_{it}$ is of dimension p and excludes time-invariant regressors for identification purposes; $\boldsymbol{\beta}(\cdot)$ is a $p \times 1$ vector of the corresponding unknown functional coefficients which are assumed to be smooth functions; and the vector of smoothing variables $\mathbf{z}_{it}$ is of dimension q. The remaining variables in the equation are all scalars. We assume the following about unobservables in the model. The individual-specific effects $\{\mu_i\}$ are an *i.i.d.* sequence (over i) with a zero mean and finite variance. Analogously, the time-specific effect λ_t is *i.i.d.* over t and also has a zero mean and finite variance. We let both μ_i and λ_t correlate with $\mathbf{x}_{it}$ and/or $\mathbf{z}_{it}$ in an arbitrary unspecified way. Such a nonparametric treatment of both unobservable effects and their correlation with the regressors in the equation renders them "fixed." Lastly, the random disturbance u_{it} is assumed to be *i.i.d.* over both i and t with a zero mean and finite variance, and $\mathbb{E}[u_{it}|\mathbf{w}_i] = 0$ with $\mathbf{w}_i = (\mu_i, \lambda_t, \mathbf{x}_{i1}', \ldots, \mathbf{x}_{iT}', \mathbf{z}_{i1}', \ldots, \mathbf{z}_{iT}')'$. Thus, eq. (4.1) is a fixed-effects model with strictly exogenous regressors.

The salient feature of our model is that it accommodates heterogeneity in *both* dimensions of the data: cross-section (μ_i) and time (λ_t). In that, it nests the more traditional functional-coefficient one-way fixed-effects panel data model with individual effects only. Namely, when $\lambda_t = 0$ for all t, model (4.1) reduces to the more standard specification first studied by [1].

The consistent estimation of (4.1) is complicated by the presence of unobservable $\{\mu_i\}$ and $\{\lambda_t\}$ which cannot be ignored due to their correlation with the regressors. A popular approach to tackling fixed effects is to transform the model to rid it of the former. For instance, this is the route taken by [24, 25] in their estimation of the one-way model whereby they remove individual effects via cross-time first-difference/within transformation *before* taking a local approximation of the transformed equation. Such an approach is non-trivial in the case with two fixed effects because (i) if first-differencing, it is not obvious what cross-sections to difference λ_t over given that there is typically no natural ordering across units and (ii) if within-transforming, the local approximation of the transformed two-way model would need to be TNq-variate which would deliver an exceptionally poor convergence rate and is practically prohibitive.[2] We therefore proceed with the alternative method whereby both fixed effects are removed *after* taking the local approximation of (4.1).

Building on Sun's (2009) idea, we generalize their smoothed LSDV approach to tackling *more than one* fixed effect in the functional-coefficient model. This extension is rather natural. Writing the model in (4.1) in the matrix from, we have

$$\mathbf{y} = \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(\mathbf{Z})\} + \mathbf{D}\boldsymbol{\mu} + \mathbf{P}\boldsymbol{\lambda} + \mathbf{u}, \tag{4.2}$$

where $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_n')'$ and $\mathbf{u} = (\mathbf{u}_1', \ldots, \mathbf{u}_n')'$ are $(nT) \times 1$ vectors, with $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ and $\mathbf{u}_i = (u_{i1}, \ldots, u_{iT})'$. The operator $\mathbb{M}\{\cdot\}$ stacks $\mathbf{x}_{it}'\boldsymbol{\beta}(\mathbf{z}_{it})$ into a $nT \times 1$ vector with the (i, t) subscripts matching those of $\mathbf{y}$ and $\mathbf{u}$. Further, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ is an $n \times 1$ vector of

[2] In case of the two-way first-differencing, the local approximation would need to be $4q$-variate which is also likely practically prohibitive.

individual fixed effects, and $\mathbf{D} = \mathbf{I}_n \otimes \mathbf{i}_T$ is an $(nT) \times n$ design matrix, where $\mathbf{I}_n$ is an identity matrix of dimension n and $\mathbf{i}_T$ denotes an $T \times 1$ vector of ones. Analogously, $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_T]'$ is a $T \times 1$ vector of time fixed effects, with $\mathbf{P} = \mathbf{i}_n \otimes \mathbf{I}_T$.

As the name suggests, the smoothed LSDV approach involves local smoothing around $\mathbf{z}_{it} = z$. First, define a $nT \times nT$ diagonal matrix of local kernel weights $\mathbf{W}_H(z) = \text{diag}\{\mathbf{K}_H(\mathbf{z}_1, z), \cdots, \mathbf{K}_H(\mathbf{z}$ with $\mathbf{K}_H(\mathbf{z}_i, z) = \text{diag}\{k_H(z_{i1}, z), \cdots, k_H(z_{iT}, z)\}$ being a $T \times T$ diagonal matrix for each i, where $k_H(\mathbf{z}_{it}, z) = k\{\mathbf{H}^{-1}(\mathbf{z}_{it} - z)\}$ is the q-variate product kernel and $\mathbf{H} = \text{diag}\{h_1, \ldots, h_q\}$ is a diagonal bandwidth matrix of dimension q. To derive the estimator for unknown functional coefficients, we then solve the following locally weighted least-squares problem:

$$\min_{\boldsymbol{\beta}(z), \boldsymbol{\mu}, \boldsymbol{\lambda}} [\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\} - \mathbf{D}\boldsymbol{\mu} - \mathbf{P}\boldsymbol{\lambda}]' \mathbf{W}_H(z) [\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\} - \mathbf{D}\boldsymbol{\mu} - \mathbf{P}\boldsymbol{\lambda}]. \tag{4.3}$$

The first-order conditions of (4.3) with respect to both fixed effects $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are

$$\mathbf{D}' \mathbf{W}_H(z) [\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\} - \mathbf{D}\boldsymbol{\mu} - \mathbf{P}\boldsymbol{\lambda}] = 0 \tag{4.4}$$

$$\mathbf{P}' \mathbf{W}_H(z) [\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\} - \mathbf{D}\boldsymbol{\mu} - \mathbf{P}\boldsymbol{\lambda}] = 0, \tag{4.5}$$

which can be solved for $\widehat{\boldsymbol{\mu}}$, i.e.,

$$\begin{aligned} \widehat{\boldsymbol{\mu}}(z) = & \left[\mathbf{I}_n - (\mathbf{D}'\mathbf{W}_H(z)\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}_H(z)\mathbf{P}(\mathbf{P}'\mathbf{W}_H(z)\mathbf{P})^{-1}\mathbf{P}'\mathbf{W}_H(z)\mathbf{D}\right]^{-1} (\mathbf{D}'\mathbf{W}_H(z)\mathbf{D})^{-1} \times \\ & \mathbf{D}'\mathbf{W}_H(z) \left[\mathbf{I}_{nT} - \mathbf{P}(\mathbf{P}'\mathbf{W}_H(z)\mathbf{P})^{-1}\mathbf{P}'\mathbf{W}_H(z)\right] (\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\}), \end{aligned} \tag{4.6}$$

and for $\widehat{\boldsymbol{\lambda}}$:

$$\begin{aligned} \widehat{\boldsymbol{\lambda}}(z) = & \left[\mathbf{I}_T - (\mathbf{P}'\mathbf{W}_H(z)\mathbf{P})^{-1}\mathbf{P}'\mathbf{W}_H(z)\mathbf{D}(\mathbf{D}'\mathbf{W}_H(z)\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}_H(z)\mathbf{P}\right]^{-1} (\mathbf{P}'\mathbf{W}_H(z)\mathbf{P})^{-1} \times \\ & \mathbf{P}'\mathbf{W}_H(z) \left[\mathbf{I}_{nT} - \mathbf{D}(\mathbf{D}'\mathbf{W}_H(z)\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}_H(z)\right] (\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\}). \end{aligned} \tag{4.7}$$

Define the two *local* within transformation matrices: $\mathbf{N}_H(z) \equiv \mathbf{I}_{nT} - \mathbf{P}[\mathbf{P}'\mathbf{W}_H(z)\mathbf{P}]^{-1}\mathbf{P}'\mathbf{W}_H(z)$ and $\mathbf{M}_H(z) = \mathbf{I}_{nT} - \mathbf{D}[\mathbf{D}'\mathbf{\Omega}_H(z)\mathbf{D}]^{-1}\mathbf{D}'\mathbf{\Omega}_H(z)$, where $\mathbf{\Omega}_H(z) \equiv \mathbf{N}'_H(z)\mathbf{W}_H(z)\mathbf{N}_H(z)$. Then, substituting $\widehat{\boldsymbol{\mu}}(z)$ and $\widehat{\boldsymbol{\lambda}}(z)$ from (4.6)–(4.7) for $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, respectively, in the objective function in (4.3) yields a concentrated locally weighted least-squares problem from which both unknown fixed effects are removed:

$$\min_{\boldsymbol{\beta}(z)} \left(\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\}\right)' \mathbf{\Sigma}_H(z) \left(\mathbf{y} - \mathbb{M}\{\mathbf{X}, \boldsymbol{\beta}(z)\}\right), \tag{4.8}$$

where $\mathbf{\Sigma}_H(z) \equiv \mathbf{M}_H(z)'\mathbf{\Omega}_H(z)\mathbf{M}_H(z)$. The riddance of fixed effects from the model is ensured by $\mathbf{N}_H(z)\mathbf{P}\boldsymbol{\lambda} = \mathbf{0}_{nT\times 1}$ and $\mathbf{M}_H(z)\mathbf{D}\boldsymbol{\mu} = \mathbf{0}_{nT\times 1}$ for all z. Upon a close examination of the local weighting matrix $\mathbf{\Sigma}_H(z)$, it is evident that $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are removed via local z-specific kernel-weighted two-way within transformation. That is, the locally approximated model is transformed by subtracting the smoothed (around $\mathbf{z}_{it} = z$) version of cross-time and cross-individual averages and adding the smoothed pooled (grand) average. Note that the two-way within transformation implied by the smoothed LSDV procedure is performed *after* the local approximation is taken thereby removing μ_i and λ_t "asymptotically" as opposed to purging them from (4.1) by directly within-transforming the model using unsmoothed (global) averages before deriving the estimator for functional coefficients.

To operationalize the estimator of $\boldsymbol{\beta}(z)$ from the profiled problem in (4.8), we rely on local-polynomial kernel approximators. For each $s = 1, \ldots, p$, the local Taylor expansion of the unknown functional coefficient around $\mathbf{z}_{it} = z$ is

$$\beta_s(\mathbf{z}_{it}) = \beta_s(z) + (\mathbf{z}_{it} - z)'\nabla_{\mathbf{z}}\beta_s(z) + (\mathbf{z}_{it} - z)'\nabla^2_{\mathbf{z}}\beta_s(z)(\mathbf{z}_{it} - z) + \ldots, \tag{4.9}$$

where $\nabla_{\mathbf{z}}\beta_s(z) = (\partial\beta_s(z)/\partial z_{1,it}, \ldots, \partial\beta_s(z)/\partial z_{q,it})'$ is a $q \times 1$ vector of first-order gradients and $\nabla^2_{\mathbf{z}}\beta_s(z) = \nabla_{\mathbf{z}'}(\nabla_{\mathbf{z}}\beta_s(z))$ is the $q \times q$ Hessian matrix of the second-order derivatives, etc. The first-order (local linear) approximation is arguably the most popular among practitioners

[16], and we adopt it here too and so do [1]. Thus, in what follows, we make use of $\beta_s(\mathbf{z}_{it}) \approx \beta_s(z) + (\mathbf{z}_{it} - z)\nabla_{\mathbf{z}}\beta_s(z)$ around $\mathbf{z}_{it} = z$.

Define a $(q+1) \times 1$ vector $\boldsymbol{\theta}_s(z) = (\beta_s(z), \nabla_{\mathbf{z}}\boldsymbol{\beta}_s(z)')'$ of unknown local parameters for each $s = 1, \ldots, p$. Then, the unknown $p \times (q+1)$ parameter matrix is defined as $\boldsymbol{\Theta}(z) = [\boldsymbol{\theta}_1(z) \ \ldots \ \boldsymbol{\theta}_p(z)]'$:

$$\boldsymbol{\Theta}(z) \equiv \begin{bmatrix} \boldsymbol{\theta}_1(z)' \\ \vdots \\ \boldsymbol{\theta}_p(z)' \end{bmatrix} = \begin{bmatrix} \beta_1(z) & \nabla\boldsymbol{\beta}_1(z)' \\ \vdots & \vdots \\ \beta_p(z) & \nabla\boldsymbol{\beta}_p(z)' \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}(z) & \nabla\boldsymbol{\beta}(z)' \end{bmatrix},$$

where the first column of the above matrix is $\boldsymbol{\beta}(\cdot)$ evaluated at z which is of primary interest. Next, define a $(q+1) \times 1$ vector of deviations from z, i.e., $\mathcal{Z}_{it}(z) = (1, (\mathbf{z}_{it} - z)')'$. For $\mathbf{z}_{it}$ close to z, we replace $\boldsymbol{\beta}(\mathbf{z}_{it})$ in (4.8) with $\boldsymbol{\Theta}(z)\mathcal{Z}_{it}(z)$ and obtain the local-linear estimator of functional coefficients $\boldsymbol{\beta}(z)$ from the locally concentrated minimization problem:

$$\min_{\boldsymbol{\Theta}(z)} (\mathbf{y} - \mathcal{X}(z)\text{vec}\{\boldsymbol{\Theta}(z)\})'\boldsymbol{\Sigma}_H(z)\left(\mathbf{y} - \mathcal{X}(z)\text{vec}\{\boldsymbol{\Theta}(z)\}\right), \tag{4.10}$$

where we stack (by columns) the unknown parameter matrix $\boldsymbol{\Theta}(z)$ into a $p(q+1) \times 1$ vector denoted by the operator vec$\{\cdot\}$, and $\mathcal{X}(z) = (\mathcal{X}_i'(z), \ldots, \mathcal{X}_n'(z))'$ is an $nT \times p(q+1)$ data matrix, with each $T \times p(q+1)$ block given by

$$\mathcal{X}_i(z) = \begin{bmatrix} \mathcal{Z}_{i1}'(z) \otimes \mathbf{x}_{i1}' \\ \vdots \\ \mathcal{Z}_{iT}'(z) \otimes \mathbf{x}_{iT}' \end{bmatrix}.$$

Lastly, solving the first-order condition of (4.10) for the unknown $\boldsymbol{\Theta}(z)$ yields the following local-linear two-way fixed-effects estimator:

$$\text{vec}\left\{\widehat{\boldsymbol{\Theta}}(z)\right\} = \left(\mathcal{X}(z)'\boldsymbol{\Sigma}_H(z)\mathcal{X}(z)\right)^{-1}\mathcal{X}(z)'\boldsymbol{\Sigma}_H(z)\mathbf{y}. \tag{4.11}$$

A few remarks are in order. Since we assume that all elements in $\mathbf{x}_{it}$ are time-varying, the identification of our fixed-effects model does not require additional restrictions on μ_i and λ_t such as the popular "zero sum" normalization à la [29] or [12] oftentimes imposed in semiparametric fixed-effects models.[3] The latter is normally necessary if the model admits time-invariant regressors, although one can sometimes achieve identification even without it: e.g., [1] identify their original one-way fixed-effects model that permits one time-invariant regressor by relying on the assumption that fixed effects are an i.i.d. sequence of random variables with a zero mean and finite variance. Similar arguments are used by [10] for a fully nonparametric one-way fixed-effects model. However, even if the identification can be achieved without restricting fixed effects, to make the local-polynomial smoothed LSDV *estimation* of the functional coefficient on the time-invariant regressor (essentially, a non-parametric intercept function) feasible, one has to restrict the unobserved effects nonetheless. This happens because the within transformation is meant to remove *any* time-invariant term. The latter is the reason why Sun's (2009) one-way estimator is operationalized using the restricted matrix $\mathbf{D}$ under the zero-sum normalization of fixed effects. In our case however, we derive the estimator under no such restriction. Having said that, should one be interested in allowing for a time-invariant x in our two-way model, the estimator in (4.11) may be made feasible by replacing $\mathbf{D}$ and $\mathbf{P}$ with their restricted counterparts $\mathbf{D}_R = [-\mathbf{i}_{n-1}\mathbf{I}_{n-1}]' \otimes \mathbf{i}_m$ and $\mathbf{P}_{\mathbf{R}} = \mathbf{i}_n \otimes [-\mathbf{i}_{m-1}\mathbf{I}_{m-1}]'$ (under $\sum_{i=1}\mu_i = 0$ and $\sum_{t=1}\lambda_t = 0$) and accordingly redefining matrices $\mathbf{N}_H(z)$, $\mathbf{M}_H(z)$, $\mathbf{\Omega}_H(z)$ and $\mathbf{\Sigma}_H(z)$.

4.3 Simulation Study

We examine the finite-sample performance of the proposed estimator in (4.11) in a series of Monte Carlo experiments. All data generating processes follow model (4.1).

We begin with the DGP with $p = q = 1$ whereby $y_{it} = x_{it}\beta(z_{it}) + \mu_i + \lambda_t + u_{it}$, where the variables are drawn as follows: $z_{it} = 0.5(\omega_{it} + \omega_{it-1})$, where $\omega_{it} \sim$ i.i.d. $\mathcal{U}(0, 0.5\pi)$;

[3] The assumption of no time-invariant regressors in functional-coefficient fixed-effects models is not without precedent, e.g., see [24].

Table 4.1. Simulation results for the two-way SLSDV estimator ($p = 1$, $q = 1$)

	$T = 3$			$T = 5$		
	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
			Case 1: x and z are correlated			
RMSE	0.1763	0.1398	0.1063	0.1233	0.0981	0.0747
MAE	0.1237	0.0967	0.0713	0.0836	0.0637	0.0478
			Case 2: x and z are uncorrelated			
RMSE	0.1850	0.1459	0.1096	0.1274	0.1011	0.0788
MAE	0.1267	0.0989	0.0723	0.0849	0.0647	0.0488

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$.

$x_{it} = 0.5(bz_{it} + x_{it-1}) + \zeta_{it}$, where $\zeta_{it} \sim$ i.i.d. $\mathcal{N}(0, 1)$; and $u_{it} \sim$ i.i.d. $\mathcal{N}(0, 0.5)$. We consider two cases: (1) x_{it} and z_{it} are correlated with $b = 1$ and (2) x_{it} and z_{it} are uncorrelated with $b = 0$. The outcome is generated with the following specification of individual and time effects: $\mu_i = c_1(\overline{z}_i + \overline{x}_i) + \rho_i$ with $\rho_i \sim$ i.i.d. $\mathcal{N}(0, 0.5)$, and $\lambda_t = c_2(\overline{z}_t + \overline{x}_t) + \varrho_t$ with $\varrho_t \sim$ i.i.d. $\mathcal{N}(0, 0.5)$, where c_1 and c_2 control the degree of correlation with regressors. Here, we set $c_1 = c_2 = 0.5$ for "fixed" effects. The functional coefficient is specified as $\beta(z_{it}) = \sin(\pi z_{it})$.

We consider cross-sectional sample sizes $n = \{50,\ 100,\ 200\}$ with the number of time periods $T = \{3,\ 5\}$. For each (n, T), we simulate the model 500 times. We use the popular [9] rule-of-thumb bandwidth for the smoothing variables. The kernel function of choice is second-order Gaussian. For each simulation, we compute the average (over z_{it}) root mean squared error (RMSE) and the average (over z_{it}) mean absolute error (MAE) for each functional coefficient function and then report their respective averages computed over 500 simulations in Table 4.1.

The results in Table 4.1 are encouraging and indicate that, in both cases, the estimation of $\beta(\cdot)$ becomes more stable as the sample size increases for both fixed T and fixed n. Both the RMSE and MAE decline significantly.

Next, we examine the performance of our estimator in higher-dimensional models. First, we consider the fixed-effects DGP with $p = 2$ regressors and $q = 1$ smoothing variable: $y_{it} =$

Table 4.2. Simulation results for the two-way SLSDV estimator ($p = 2$, $q = 1$)

		$n = 50$	$n = 100$	$n = 200$
$\beta_1(\cdot)$	RMSE	0.1798	0.1427	0.1046
	MAE	0.1202	0.0933	0.0685
$\beta_2(\cdot)$	RMSE	0.1900	0.1489	0.1129
	MAE	0.1301	0.1015	0.0759

Reported are the results for the functional coefficient estimators $\widehat{\beta}_1(\cdot)$ and $\widehat{\beta}_2(\cdot)$. $T = 3$ throughout.

Table 4.3. Simulation results for the two-way SLSDV estimator ($p = 1$, $q = 2$)

	$n = 50$	$n = 100$	$n = 200$
RMSE	0.3153	0.2592	0.2160
MAE	0.2047	0.1659	0.1351

Reported are the results for the functional coefficient estimator $\widehat{\beta}_1(\cdot)$. $T = 3$ throughout.

$x_{1,it}\beta_1(z_{it})+x_{2,it}\beta_2(z_{it})+\mu_i+\lambda_t+u_{it}$, where $z_{it} = 0.5(\omega_{it}+\omega_{it-1})$; $x_{s,it} = 0.5(z_{it}+x_{s,it-1})+\zeta_{s,it}$ for $s = 1, \dots, p$; $\mu_i = c_1(\overline{z}_i + 0.5\overline{x}_{1,i} + 0.5\overline{x}_{2,i}) + \rho_i$ and $\lambda_t = c_2(\overline{z}_t + 0.5\overline{x}_{1,t} + 0.5\overline{x}_{2,t}) + \varrho_t$ with $c_1 = c_2 = 0.5$. The random terms ω_{it}, $\zeta_{s,it}$, u_{it}, ρ_i and ϱ_t are drawn as before. The functional coefficients are specified as $\beta_1(z_{it}) = 1 + z_{it}^3/3$ and $\beta_2(z_{it}) = \sin(\pi z_{it})$. Table 4.2 summarizes these results. Second, we increase the number of variables that enter the model nonparametrically. The fixed-effects DGP with $p = 1$ regressor but $q = 2$ smoothing variables is $y_{it} = x_{it}\beta(z_{it}) + \mu_i + \lambda_t + u_{it}$, where $z_{l,it} = 0.5(\omega_{l,it} + \omega_{l,it-1})$ for $l = 1, \dots, q$; $x_{it} = 0.5(z_{q,it}+x_{it-1})+\zeta_{it}$; $\mu_i = c_1(0.5\overline{z}_{1,i}+0.5\overline{z}_{2,i}+\overline{x}_i)+\rho_i$ and $\lambda_t = c_2(0.5\overline{z}_{1,t}+0.5\overline{z}_{2,t}+\overline{x}_t)+\varrho_t$ with $c_1 = c_2 = 0.5$. The remaining random terms $\omega_{l,it}$, ζ_{it}, u_{it}, ρ_i and ϱ_t are drawn as before. The functional coefficient is specified as $\beta(z_{it}) = 1+z_{1,it}z_{2,it}+z_{2,it}^2$. The corresponding results are reported in Table 4.3.

From Tables 4.2–4.3, we see that the estimator continues to perform well when the dimensionality of the model rises. As expected of a consistent estimator, RMSE declines with the increase in n.[4]

[4]This chapter includes excerpts from " Halder, S., and Malikov, E. Smoothed LSDV Estimation of Functional-Coefficient Panel Data Models with Two-Way Fixed Effects, Economics Letters. 2020, Volume 192, Article 109239."

Chapter 5

Empirical Applications

We showcase the practical usefulness of our two-way SLSDV estimator in two different scenarios both for balanced dataset as well as for unbalanced dataset.

5.1 Balanced Data

A balanced panel dataset is a dataset in which each panel member is observed every year. Consequently, if a balanced panel contains n panels and T periods, the number of total observations in the dataset is $N = n \times T$.[1]

We showcase the practical usefulness of our two-way SLSDV estimator by revisiting the estimation of the so-called environmental Kuznets curve (EKC) that relates environmental quality to economic development with the focus on (the oftentimes overlooked) temporal variability in the coefficients. The empirical literature on the EKC hypothesis is broad; e.g., see [6] and [4] for excellent surveys. The gist of such studies essentially boils down to the estimation of a "reduced-form" regression of a pollution variable on the region's per-capita income, traditionally, using panel data that permit controlling for unobservable confounders via fixed effects. Most prevalently, the adopted parametric specification is cubic (seldom quadratic) arguably due to its ability to fit different relationships including the inverted-U shape implied by the hypothesis e.g.,to name a few [5, 7, 4, 8], with the usual model taking the following parametric form:

$$P_{it} = \mu_i + \lambda_t + \beta_1 Y_{it} + \beta_2 Y_{it}^2 + \beta_3 Y_{it}^3 + u_{it}, \tag{5.1}$$

[1]We are thankful to Dan Millimet and Alfonso Flores-Lagunes for sharing their data for this empirical application

where P_{it} and Y_{it} are, respectively, the pollutant emissions and income per capita in region i at time t; The two-way region- and time-specific unobservables are treated as being "fixed" effects.

Surprisingly, despite that the so-called "technique effect" capturing improvements in technology and efficiency over time is one of the key proximate factors normally invoked to explain the EKC hypothesis (and the shape of the curve it predicts), the estimated EKC regressions just like the one in (5.1) normally assume time-invariance of the relationship between P and Y. But it is only natural to expect the pollution-income nexus to also evolve over time with technological change. In this paper, we seek to examine the appropriateness of this implicit time-invariance assumption by relaxing parameter constancy in (5.1) to let the slope coefficients smoothly vary with time in an unspecified way. By letting the coefficients vary with time in an arbitrary nonparametric fashion, we are able to accommodate potential temporal variation in the relationship between pollution and income in a flexible way that is robust to misspecification.[2] Thus, our preferred EKC regression is a semiparametric functional-coefficient two-way fixed-effects model:

$$P_{it} = \mu_i + \lambda_t + \beta_1(D_t)Y_{it} + \beta_2(D_t)Y_{it}^2 + \beta_3(D_t)Y_{it}^3 + u_{it}, \tag{5.2}$$

where the functional coefficient vector $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \boldsymbol{\beta}_2(\cdot)', \boldsymbol{\beta}_3(\cdot)')'$ is a function of a (scalar) ordered discrete time variable D_t taking on T different values in $\{1, 2, \ldots, T\}$. Since the smoothing variable is not continuous, equation (5.2) is estimated via the local-*constant* version of our proposed two-way SLSDV estimator using the cross-validated bandwidth and [2]'s (2007) kernel function for ordered discrete variables.

By virtue of treating time as an *ordered* "factor" variable entering nonparametric functional coefficients, we essentially achieve the estimation of a fully saturated flexible model (with a complete set of time dummies) without the need to actually cell-split the sample

[2] [7] also discuss the important of accounting for temporal variation in the standard EKC regressions driven by technological changes but do not go beyond the inclusion of time trends, thereby restrictively assuming additivity while still maintaining the time-invariance of slopes.

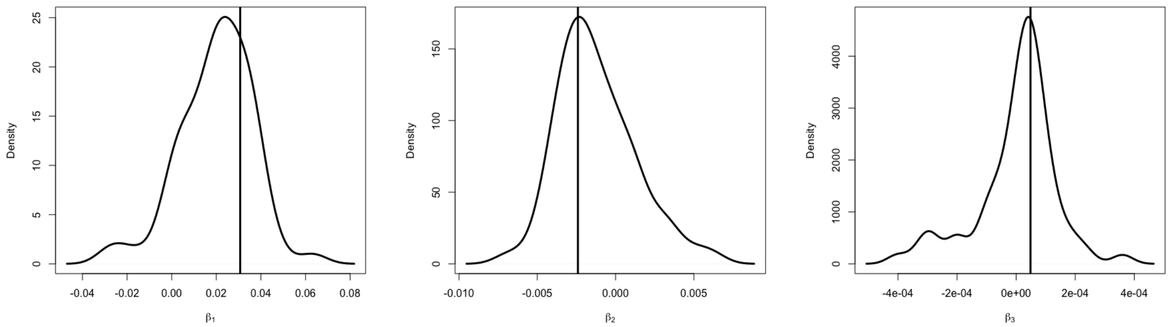

Figure 5.1. Distribution of the semiparametric EKC coefficient estimates over the years for NO_x

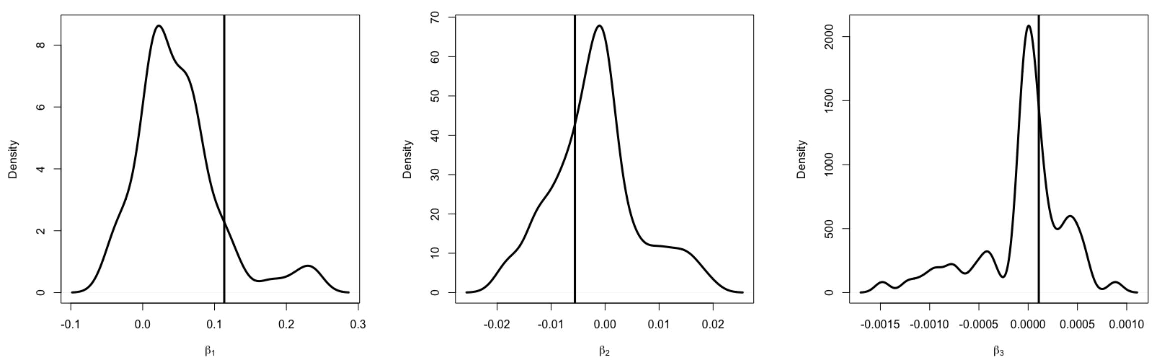

Figure 5.2. Distribution of the semiparametric EKC coefficient estimates over the years for SO_2

based on years. By relying on recent advancements in the kernel estimation of functional-coefficient models with exclusively discrete smoothing variables [see[3]], we circumvent the need for time dummies and, further, can use relevant information from "similar" time periods (from before and after) during estimation of the coefficients for each time period by "smoothing" them over time. Not least importantly, such a kernel-based nonparametric treatment of a scalar time variable comes at no cost in terms of the speed of convergence since smoothing over categorical variables does not contribute to the "curse of dimensionality." In the absence of continuous smoothing variables (like in our case), the estimator of unknown coefficients exhibits a parametric rate.

We estimate the EKC at the U.S. state level. The pollutants of focus are NO_x and SO_2. The annual panel contains observations for $n = 48$ contiguous states during 1929–1994 ($T = 66$). The state-level per-capita emissions P_{it} and per-capita income Y_{it} are in thousands short tons and thousands of 1987 U.S. dollars, respectively. See [26] and [8] for summary statistics and further details about the data.

Figures 5.1–5.2 summarize year-specific estimates of the three functional coefficients from the semiparametric EKC in (5.2) in the form of kernel densities plotted together with their fixed-coefficient counterparts (vertical lines) from a fully parametric model in (5.1). For both pollutants, the data point to non-negligible temporal heterogeneity in the EKC relationship. In most cases, the parametric coefficient estimates are not even near the mode of the corresponding semiparametric estimates. To see the latter more clearly, we plot the estimated EKC using both models.

Since our semiparametric specification produces time-varying coefficients, the implied EKC relationship is also time-varying. To avoid clutter of 66 year-specific curves, we instead split the sample period into decades and plot the decade-specific EKCs using medians of year-specific functional coefficients from each subperiod. These are 1929–1939, 1940–1949, 1950–1959, ..., 1980–1989, 1990–1994, with the first/last subperiods being longer/shorter than others given the endpoints of our sample period. Each curve is plotted in the range of Y_{it} values observed during the corresponding decade. The semiparametric decade-specific EKCs are plotted along with that estimated using a fully parametric model, which is "global" over the entire sample period, in Figures 5.3 and 5.5. These figures show that, in the face of temporal changes, a time-invariant fixed-coefficient model can produce an incomplete, if not distorted, picture of the EKC relationship.

To begin with, consider the case of NO_x pollutants. While both models produce an inverted-U-shaped relationship consistent with the EKC hypothesis, our semiparametric time-varying model produces additional important insights into the pollution-income nexus that the more traditional fixed-coefficient model cannot deliver by design. The latter includes not only the understanding of the "drift" of the relationship over time as per-capita incomes of the states grow but, perhaps more importantly, the evolution of the turning point with technological changes and the states' position on the EKC relative thereto. The fixed-coefficient model estimates the "global" turning point around the quite optimistic $8,700. Although this is near the median of year-specific estimates of the turning point implied by our time-varying model ($8,600), our model also indicate a considerable variability in the turning point over the years, with the post-1967 period characterized by significantly larger values of the point (see Table 5.1 in the Online Appendix). Figure 5.4 plots the time evolution of the semiparametric NO_x turning point estimates, from where it is evident that the hump of the EKC curve was generally drifting to the right (higher income values) for most years until the reversal of this trend in the last ten years or so.

The contrast between the two models is even more stark in the case of SO_2 emissions. Consistent with the results reported by [26] and [8], the parametric specification produces a monotonically increasing EKC which offers no support for the hypothesis. The results from the time-varying model however lead to a different conclusion. From Figure 5.5, the estimated EKC is *non*-monotonic in the earlier years before the 1970s, which also can be concluded from Figure 5.6 that presents yearly estimates of the turning point for SO_2 (also see Table 5.2). The semiparametric estimates suggest the relative stability of the turning point around the overall median of $8,600 until around 1970, which was only then followed by the period characterized by dramatically higher or even missing[3] values of the turning point. Thus, the lack of empirical evidence in support of the EKC hypothesis produced by the time-invariant model is likely driven by the last third of the sample period only. This highlights the practical importance of using more flexible specifications that allow the EKC relationship to evolve over time.

[3]Missing due to positive monotonicity of the EKC in some years.

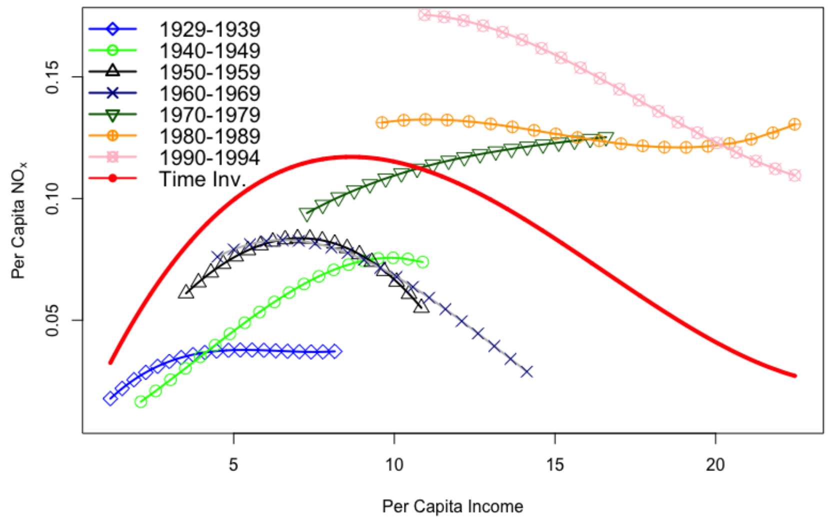

Figure 5.3. Semiparametric estimates of the EKC for NO_x across decades (decade-specific curves are plotted in the range of per-capita income values observed during the corresponding decade; the solid line is the EKC from a time-invariant parametric model)

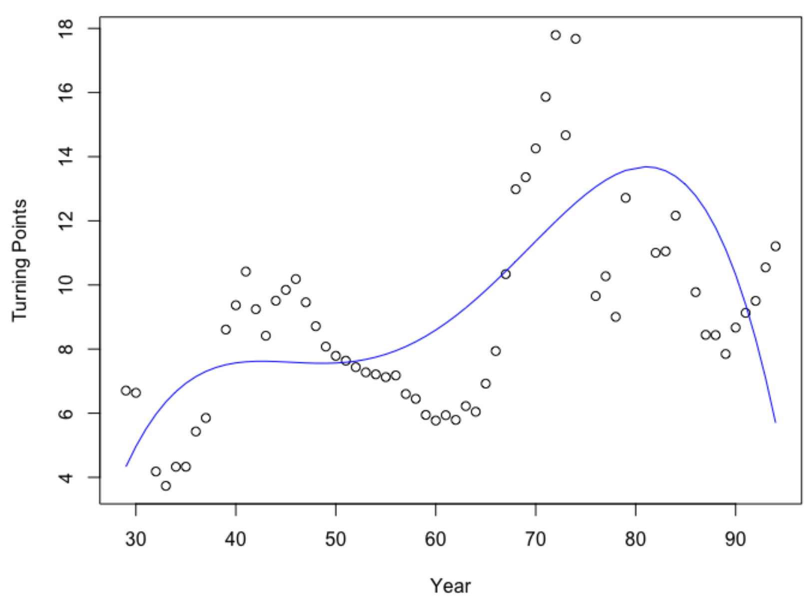

Figure 5.4. Evolution of the turning point estimates over the years for NO_x (solid line is the fitted fourth-degree polynomial)

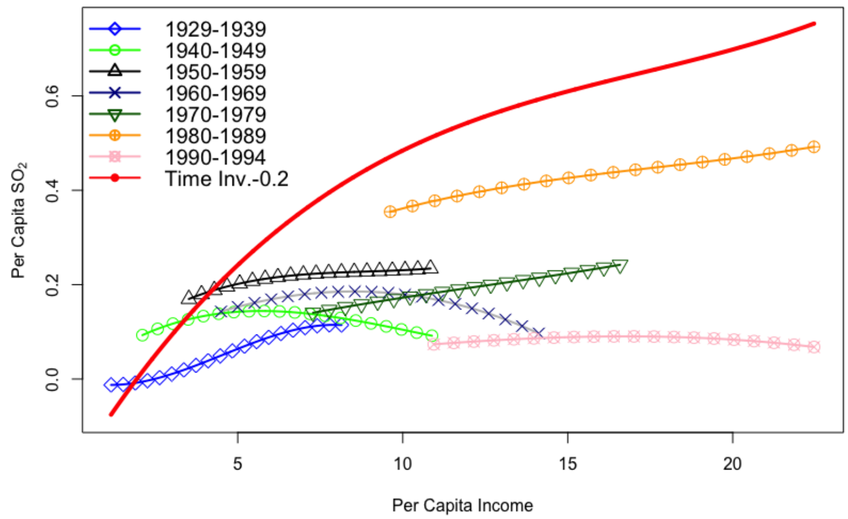

Figure 5.5. Semiparametric estimates of the EKC for SO_2 across decades (decade-specific curves are plotted in the range of per-capita income values observed during the corresponding decade; the solid line is the EKC from a time-invariant parametric model, shifted down by 0.2)

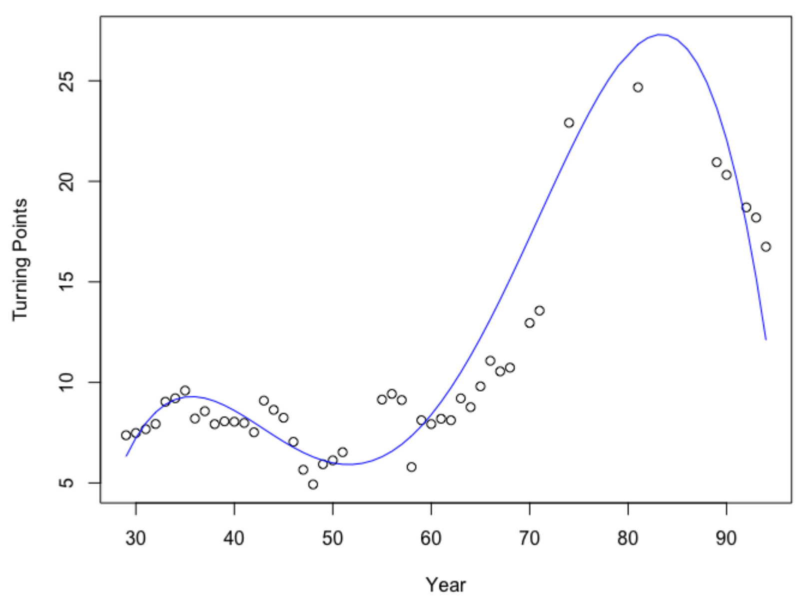

Figure 5.6. Evolution of the turning point estimates over the years for SO_2 (solid line is the fitted fourth-degree polynomial)

Besides directly comparing our semiparametric EKC estimates to those from a fully parametric fixed-coefficient model, we also formally discriminate between the two specifications. Namely, we test the null hypothesis of a parametric time-invariant fixed-coefficient model (5.1) against our semiparametric alternative in (5.2). This is, essentially, the test of overall relevancy of D_t, or parameter constancy. To test this hypothesis, we use [22]'s nonparametric goodness-of-fit test, or the "nonparametric F-test," based on the comparison of the restricted and unrestricted models with the corresponding residual-based test statistic given by $T_n = (RSS_0 - RSS_1)/RSS_1$, where RSS_0 and RSS_1 are the residual sums of squares under the (restricted parametric) null and the (unrestricted semiparametric) alternative, respectively. Intuitively, the test statistic is expected to converge to zero under the null and is positive under the alternative; hence the test is one-sided. To approximate the null distribution of T_n, we use wild panel-data block-bootstrap by resampling residuals from the restricted model.

Steps for Block Bootstrap:

After repeating B bootstrap estimations of $[\beta_1(D_t), \beta_2(D_t), \beta_3(D_t)]$, use the empirical distribution of these B bootstrap estimates of parameter functions $\left\{ \left[\widehat{\beta}_1^b(D_t), \widehat{\beta}_2^b(D_t), \widehat{\beta}_3^b(D_t) \right]; b = 1, \ldots, B \right\}$ to construct bias-corrected confidence intervals corresponding to each $\widehat{\beta}_j(D_t)$ estimate for $j = 1, 2, 3$.

Namely, we estimate $(1 - a)$ 100% confidence bounds for each $\widehat{\beta}_j(D_t)$ as intervals between the $[a_1 \times 100]$th and $[a_2 \times 100]$th percentiles of the corresponding bootstrap distribution $\left\{ \widehat{\beta}_j^b(D_t); b = 1, \ldots, B \right\}$ with

$$a_1 = \Phi \left(2\widehat{\Phi}_0 + Z_{\alpha/2} \right) \tag{5.3}$$

$$a_2 = \Phi \left(2\widehat{\Phi}_0 + Z_{(1-\alpha/2)} \right) \tag{5.4}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, Z_α is the $(a \times 100)$th percentile of the standard normal distribution, and

$$\widehat{\Phi}_0 = \Phi^{-1}\left(\#\left\{\widehat{\beta}_j^b(D_t) \leq \widehat{\beta}_j(D_t)\right\}/B\right), \tag{5.5}$$

where $\#\{A\}$ is the count of A being true.

Bias-corrected confidence intervals for the turning points.: For each bootstrap iteration b and each D_t value within, use $\left[\widehat{\beta}_1^b(D_t), \widehat{\beta}_2^b(D_t), \widehat{\beta}_3^b(D_t)\right]$ to compute the bootstrap estimates of the turning point, say, $\widehat{\alpha}^b(D_t)$.

Then, for each value of D_t, use the empirical distribution of B bootstrap estimates of $\left\{\widehat{\alpha}^b(D_t); b = 1, \ldots, B\right\}$ to construct bias-corrected confidence intervals corresponding to each $\widehat{\alpha}(D_t)$ computed using the original $\left[\widehat{\beta}_1(D_t), \widehat{\beta}_2(D_t), \widehat{\beta}_3(D_t)\right]$.

Namely, we estimate $(1-a)\,100\%$ confidence bounds for each $\widehat{\alpha}(D_t)$ as intervals between the $[a_1 \times 100]$th and $[a_2 \times 100]$th percentiles of the bootstrap distribution $\left\{\widehat{\alpha}^b(D_t); b = 1, \ldots, B\right\}$ with

$$a_1 = \Phi\left(2\widehat{\Phi}_0 + Z_{\alpha/2}\right) \tag{5.6}$$

$$a_2 = \Phi\left(2\widehat{\Phi}_0 + Z_{(1-\alpha/2)}\right) \tag{5.7}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, Z_α is the $(a \times 100)$th percentile of the standard normal distribution, and

$$\widehat{\Phi}_0 = \Phi^{-1}\left(\#\left\{\widehat{\alpha}^b(D_t) \leq \widehat{\alpha}(D_t)\right\}/B\right), \tag{5.8}$$

where $\#\{A\}$ is the count of A being true.

Specification Test:

To formally discriminate our model against the fixed-coefficient alternative specification, we use Ullah's (1985) nonparametric goodness-of-fit test. Specifically, we are interested in testing the null hypothesis of a fully parametric fixed-coefficient model:

$$\mathrm{H}_0: \quad E\left[P_{it}|\mathbf{w}_i\right] = \beta_1 Y_{it} + \beta_2 Y_{it}^2 + \beta_3 Y_{it}^3 + \mu_i + \lambda_t, \tag{5.9}$$

against the alternative hypothesis (our functional-coefficient model):

$$\mathrm{H}_1: \quad E\left[P_{it}|\mathbf{w}_i\right] = \beta_1(D_t) Y_{it} + \beta_2(D_t) Y_{it}^2 + \beta_3(D_t) Y_{it}^3 + \mu_i + \lambda_t, \tag{5.10}$$

where $\mathbf{w}_i = (\mu_i, \lambda_t, Y_{it}, \ldots, Y_{it}^3)'$.

This is, essentially, the test of overall relevancy of $\mathbf{z}_{it}$, or parameter constancy. To test these hypotheses, we use a "nonparametric F-test" based on the comparison of the restricted and unrestricted models. First, let the estimator under H_0 be denoted by "tilde" whereas the estimator under H_1 be denoted by "hat." Then, the residual-based test statistic is

$$T_n = \frac{RSS_0 - RSS_1}{RSS_1}, \tag{5.11}$$

where $RSS_0 = \sum_i \sum_t \widetilde{u}_{it}^2$ and $RSS_1 = \sum_i \sum_t \widehat{u}_{it}^2$ are respectively the residual sum of squares under H_0 and H_1, with the corresponding conditional-mean residuals defined as $\widetilde{u}_{it} = P_{it} - \widetilde{E}\left[P_{it}|\mathbf{w}_i\right] = P_{it} - \widetilde{\beta}_1 Y_{it} - \widetilde{\beta}_2 Y_{it}^2 - \widetilde{\beta}_3 Y_{it}^3 - \widetilde{\mu}_i - \widetilde{\lambda}_t$ and $\widehat{u}_{it} = P_{it} - \widehat{E}\left[P_{it}|\mathbf{w}_i\right] = P_{it} - \widehat{\beta_1(D_t)} Y_{it} - \widehat{\beta_2(D_t)} Y_{it}^2 - \widehat{\beta_3(D_t)} Y_{it}^3 - \widehat{\mu}_i - \widehat{\lambda}_t$. The estimated residuals under the null can be obtained via the standard linear two-way fixed-effect estimator, with the fixed effects recovered under the $\sum_i \mu_i = 0$ and $\sum_t \lambda_t = 0$ restrictions via $\widetilde{\mu}_i = \frac{1}{T}\sum_t (P_{it} - \widetilde{\beta}_1 Y_{it} - \widetilde{\beta}_2 Y_{it}^2 - \widetilde{\beta}_3 Y_{it}^3)\ \forall i$ and $\widetilde{\lambda}_t = \frac{1}{n}\sum_i (P_{it} - \widetilde{\beta}_1 Y_{it} - \widetilde{\beta}_2 Y_{it}^2 - \widetilde{\beta}_3 Y_{it}^3)\ \forall t$.

We use bootstrap to approximate the distribution of T_n. The wild panel-data block bootstrap algorithm is as follow.

(1) Using the original data, estimate both the restricted model (under the null) and the unrestricted model (under the alternative). Obtain the corresponding residuals $\{\widetilde{u}_{it}\}$

and $\{\widehat{u}_{it}\}$. Use these to compute the test statistic T_n. Also, save the parameter estimates under H_0 $\{\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\lambda}}\}$.

(2) Generate bootstrap weights w_i^b for all $i = 1, \ldots, n$ from the two-point mass distribution:

$$w_i^b = \begin{cases} \left(1+\sqrt{5}\right)/2 & \text{with prob. } \left(\sqrt{5}-1\right)/\left(2\sqrt{5}\right) \\ \left(1-\sqrt{5}\right)/2 & \text{with prob. } \left(\sqrt{5}+1\right)/\left(2\sqrt{5}\right) \end{cases} \tag{5.12}$$

Next, for each observation (i,t) with $i = 1, \ldots, n$ and $t = 1, \ldots, T$, generate a new bootstrap disturbance using the residuals from the model under H_0: $u_{it}^b = w_i^b \times \widetilde{u}_{it}$.

(3) Construct a new bootstrap outcome variable based on the specification under H_0: $P_{it}^b = \widetilde{\beta}_1 Y_{it} + \widetilde{\beta}_2 Y_{it}^2 + \widetilde{\beta}_3 Y_{it}^3 + \widetilde{\mu}_i + \widetilde{\lambda}_t + u_{it}^b$ for all $i = 1, \ldots, n$ and $t = 1, \ldots, T$. The bootstrap sample now is given by $\{P_{i,n}^b;\ i = 1, \ldots, n\}$.

(4) Reestimate both the restricted and unrestricted models using the bootstrap sample from step (3) to obtain bootstrap residuals $\{\widetilde{u}_{i,n}^b = P_{it}^b - \widetilde{\beta}_1^{\,b} Y_{it} - \widetilde{\beta}_2^{\,b} Y_{it}^2 - \widetilde{\beta}_3^{\,b} Y_{it}^3 - \widetilde{\mu}_i^b - \widetilde{\lambda}_t^b\}$ and $\{\widehat{u}_{i,n}^b = P_{it}^b - -\widehat{\beta_1(D_t)}^{\,b} Y_{it} - \widehat{\beta_2(D_t)}^{\,b} Y_{it}^2 - \widehat{\beta_3(D_t)}^{\,b} Y_{it}^3 - \widehat{\mu}_i^b - \widehat{\lambda}_t^b\}$ under H_0 and H_1, respectively. Use these residuals, to compute the bootstrap test statistic T_n^b.

(5) Repeat steps (2)–(4) B times.

(6) Use the empirical distribution of $B+1$ bootstrap statistics $\{T_n^b\}$, where the first bootstrap test statistic equals the test statistic T_n calculated from the original data in Step 1, to obtain p-value as $\sum_b 1\left\{T_n^b \geq T_n\right\}/(B+1)$.

While we cannot reject the null of a time-invariant specification at the conventional significance level for the $\mathrm{NO_x}$ pollutant, the empirical evidence however favors our more flexible, semiparametric time-varying EKC model in the case of $\mathrm{SO_2}$, with the corresponding bootstrap p-value of 0.044. This highlights the caution practitioners ought to exercise in their choice of the constant-parameter EKC specifications.

5.2 Unbalanced Data

We also showcase the practical usefulness of our two-way SLSDV estimator using an unbalanced dataset by revisiting the role of management for production. An unbalanced panel is a dataset in which at least one panel is not observed every period. Therefore, if an unbalanced panel contains n panel and T periods, then the following strict inequality holds for the total number of observations in the dataset: $N < n \times T$.

Researchers in business economics have long argued that management is an important intangible input to production contributing significantly to productivity differences across firms and, when aggregated, across countries. However, until recently, the data on systematic measurements of management practices have been lacking. The World Management Survey developed and since expanded by [30] and their coauthors was the major step forward in this regard, and the empirical management literature has ever since been growing.

In our empirical application, we revisit the estimation of management-augmented firm production functions from [28] that are aimed to measure the association between the firm's management practices and its production performance. Specifically, our point of departure is the (fully parametric) Cobb-Douglas specification akin to their baseline fixed-effects model:

$$y_{it} = \beta_1 k_{it} + \beta_2 l_{it} + \beta_3 m_{it} + \mu_i + \lambda_t + u_{it}, \tag{5.13}$$

where y_{it} is the logged valued-added, k_{it} is the logged physical capital (fixed assets), l_{it} is the logged employment, and m_{it} is the logged management score.[4]

Our focus is on relaxing the time-invariance of production technology and, specifically, of the relationship between management and firm output implicitly assumed in (5.13). It is only natural to expect the role of management practices for firm production to change

[4]We differ from [28] regression in that we log the management variable to facilitate the elasticity-like interpretation, exclude additional firm controls such as firm age and include year effects. We do so because we do not seek to "replicate" their analysis (for one, we do not have access to their data) but rather use their research as a context for illustrating our estimator. Also, adding the controls has little implication for our main results.

over time due to both the technological and institutional changes. We therefore let the input elasticities, including that corresponding to the managerial input, smoothly vary with (discretized) time in an unspecified way. By letting the production-function coefficients vary with time in an arbitrary nonparametric fashion, we are able to accommodate potential temporal variation in the relationship between management and output in a flexible way that is robust to misspecification. Thus, our preferred production function is a semiparametric functional-coefficient two-way fixed-effects model:

$$y_{it} = \beta_1(D_t)k_{it} + \beta_2(D_t)l_{it} + \beta_3(D_t)m_{it} + \mu_i + \lambda_t + u_{it}, \tag{5.14}$$

where the functional coefficient vector $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \beta_2(\cdot)', \beta_3(\cdot)')'$ is a function of a (scalar) ordered *discrete* time variable D_t taking on T different values in $\{1, 2, \ldots, T\}$ with T being small. We opt to discretize time as opposed to treating it continuously by defining $D_t = t/T$ as commonly done in the semi/nonparametric literature on nonstationary processes e.g., [32] mainly because most empirical applications in microeconomics deal with short annual panels implying that $n \gg T$ with the measurement of time (years) being clearly discrete. Our data are exactly that, which also helps keep the illustration of our estimator as relevant for such applications as possible. For more comfort, one may choose to think of D_t not as the "time" index but, say, as a normalized proxy/measure of the global technological stock shareable by all owing to its non-rivalry and low excludability. For other examples of a categorical treatment of the smoothing time variable in the context of production function estimation, also see [31]. A discrete treatment of D_t also has the secondary benefit of having no adverse impact on the convergence rate of the estimator.

Since the smoothing variable D_t is not continuous, equation (5.14) is estimated via the local-*constant* version of our proposed two-way SLSDV estimator using the AICc-optimal bandwidth and [2] kernel function for ordered discrete variables.

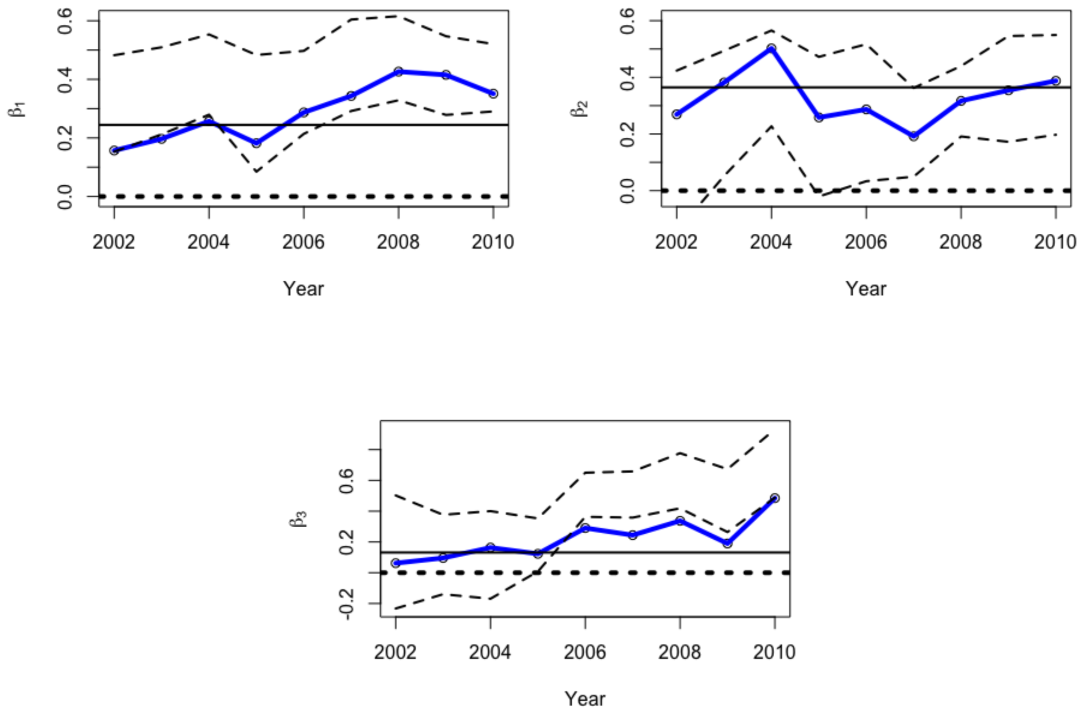

Figure 5.7. Semiparametric input elasticity estimates over the years [the management elasticity of interest $\widehat{\beta}_3(D_t)$ is in bottom plot]

We use the publicly available 2002–2010 ($T = 9$) combined survey data from across twelve countries used in [27] that contain repeatedly measured company accounts information. The sample includes $n = 1,345$ firms. The data are summarized in Table 5.3. We refer the reader to the World Management Survey website for the details on data. The output and input variables are demeaned prior to the estimation to better fit the intercept-free specification.

Figure 5.7 summarizes year-specific estimates of the three functional coefficients from the semiparametric management-augmented production function in (5.14) plotted along with their fixed-coefficient counterparts (solid horizontal lines) from a fully parametric model in (5.13) for comparison. In the figure, each point estimate of $\boldsymbol{\beta}(D_t)$ is accompanied by the corresponding 95% accelerated bias-corrected bootstrap confidence intervals (dashed lines) which are second-order accurate and provide means not only to correct for the estimator's finite-sample bias but also to account for higher-order moments (particularly, skewness) in the sampling distribution.[5]

Our interest is in the management elasticity $\beta_3(D_t)$. Overall, the data point to non-negligible temporal heterogeneity in the relationship between the firm's management practices and its production performance. The semiparametric estimates of the management elasticity are statistically significant in the year 2005 onward and suggest a growing importance of the managerial input for production over time (also see Table 5.4). Contrasting these estimates with the fixed-coefficient counterpart of the more modest effect size (0.13), we see that, in the face of technological and institutional changes, a time-invariant specification can underestimate the contribution of management to production thereby leading to incomplete, if not distorted, insights into the role of management practices for firm performance. This highlights the practical importance of using more flexible specifications.[6]

[5]Note that, given these "corrections," the point estimates may sometimes lie outside the confidence interval.

[6]This chapter includes excerpts from " Halder, S., and Malikov, E. Smoothed LSDV Estimation of Functional-Coefficient Panel Data Models with Two-Way Fixed Effects, Economics Letters. 2020, Volume 192, Article 109239."

Table 5.1. Semiparametric estimates of the EKC turning points for NO_x

Year	Turning Point	Year	Turning Point
1929	6.706*	1962	5.792*
1930	6.636*	1963	6.223*
1931	—	1964	6.046*
1932	4.182*	1965	6.924*
1933	3.735*	1966	7.939*
1934	4.330*	1967	10.337
1935	4.332*	1968	12.985*
1936	5.430*	1969	13.360*
1937	5.853	1970	14.256*
1938	—	1971	15.864
1939	8.609	1972	17.792*
1940	9.367*	1973	14.669*
1941	10.418	1974	17.676*
1942	9.244*	1975	—
1943	8.416*	1976	9.655*
1944	9.508	1977	10.271*
1945	9.847*	1978	9.003
1946	10.184*	1979	12.716*
1947	9.459	1980	87.338*
1948	8.714*	1981	—
1949	8.077*	1982	11.002*
1950	7.789	1983	11.049
1951	7.638	1984	12.161*
1952	7.436*	1985	—
1953	7.274*	1986	9.776*
1954	7.212*	1987	8.444*
1955	7.125*	1988	8.436
1956	7.179*	1989	7.847*
1957	6.600*	1990	8.670*
1958	6.449*	1991	9.129*
1959	5.947*	1992	9.503
1960	5.771*	1993	10.547*
1961	5.939*	1994	11.205*

Reported are the year-specific turning points computed using the semiparametric EKC functional coefficient estimates for each year. Values are omitted if negative or not satisfying the second-order condition for the local maximum. The asterisk signifies statistical significance at the 5% level.

Table 5.2. Semiparametric estimates of the EKC turning points for SO_2

Year	Turning Point	Year	Turning Point
1929	7.373*	1962	8.118*
1930	7.485	1963	9.207*
1931	7.676*	1964	8.768*
1932	7.933*	1965	9.805*
1933	9.038*	1966	11.076*
1934	9.211*	1967	10.549*
1935	9.594*	1968	10.733*
1936	8.204*	1969	—
1937	8.569*	1970	12.954*
1938	7.926*	1971	13.571*
1939	8.065*	1972	—
1940	8.045	1973	—
1941	7.983*	1974	22.910*
1942	7.519*	1975	—
1943	9.094*	1976	—
1944	8.636	1977	—
1945	8.242	1978	—
1946	7.044*	1979	—
1947	5.656	1980	45.021*
1948	4.926*	1981	24.677*
1949	5.930*	1982	—
1950	6.125*	1983	—
1951	6.530*	1984	—
1952	—	1985	—
1953	—	1986	—
1954	—	1987	—
1955	9.145*	1988	—
1956	9.429*	1989	20.949*
1957	9.122*	1990	20.322*
1958	5.790*	1991	—
1959	8.119*	1992	18.703*
1960	7.927*	1993	18.196*
1961	8.193*	1994	16.746

Reported are the year-specific turning points computed using the semiparametric EKC functional coefficient estimates for each year. Values are omitted if negative or not satisfying the second-order condition for the local maximum. The asterisk signifies statistical significance at the 5% level.

Table 5.3. Data summary statistics

Variable	*Mean*	*Min*	*1st Qu.*	*Median*	*3rd Qu.*	*Max*
Value Added (Y)	249,942.1	25.9	28,945.0	67,806.5	183,019.9	16,067,545.7
Capital (K)	60,588.4	2.0	4,505.9	12,530.5	37,158.9	4,266,050.4
Labor (L)	880.3	4.0	159.8	271.0	631.2	65,682.0
Management (M)	3.06	1.06	2.67	3.06	3.47	4.86

Value added and the tangible fixed assets (capital) are in USD; labor is the number of employees; management score is the average of 18 management questions.

Table 5.4. Semiparametric estimates of the management elasticity

Year	*Estimate*	*Lower Bound*	*Upper Bound*
2002	0.0621	–0.2327	0.5022
2003	0.0957	–0.1397	0.3764
2004	0.1633	–0.1690	0.4003
2005	0.1227	0.0081	0.3527
2006	0.2903	0.3632	0.6498
2007	0.2441	0.3581	0.6581
2008	0.3372	0.4184	0.7767
2009	0.1898	0.2641	0.6728
2010	0.4846	0.4876	0.9321

Bounds for the accelerated bias-corrected bootstrap intervals are at the 95% confidence.

Chapter 6

Concluding Remarks

Simulation study for our proposed modified Rodriguez- Poo & Soberon estimator performs better in finite samples in all scenarios considered. The existing semiparametric estimators for varying-coefficient fixed-effects models exclusively assume one-way fixed effects, typically in the dimension of cross-sectional units. We extend the [1] estimator to the case of a functional-coefficient model with *two*-way fixed effects whereby we allow for unobservable heterogeneity in both dimensions of the data: cross-section and time. Both fixed effects are (asymptotically) concentrated out of the model via locally smoothed two-dimensional within transformation. Simulations show that the estimator performs well in finite samples. We also showcase its practical usefulness. [1]

[1] This chapter includes excerpts from " Halder, S., and Malikov, E. Smoothed LSDV Estimation of Functional-Coefficient Panel Data Models with Two-Way Fixed Effects, Economics Letters. 2020, Volume 192, Article 109239."

Bibliography

[1] Sun, Y., Carroll, R. J., and Li, D., *Semiparametric Estimation of fixed Effects Panel Data Varying Coefficient Models*, Advances in Econometrics, **25**(2009), 101-129, .

[2] Li, Qi and Racine, Jeffrey S., *Nonparametric Econometrics: Theory and Practice*, Princeton University Press (2007).

[3] Li, Q. and Ouyang, D and Racine, J. S., *Categorical semiparametric varying-coefficient models*, Journal of Applied Econometrics, **28** (2013), 551-579.

[4] He, J., *is the environmental Kuznets curve hypothesis valid for developing countries?* A survey, *Working Paper*, Université de Sherbrooke (2007).

[5] Grossman, G. M. and Krueger, A. B., *Economic growth and the environment*, Quarterly Journal of Econometrics, **110** (1995), no. 2, 353–377.

[6] Dinda, S., *Environmental Kuznets Curve hypothesis: A survey*, Ecological Economics, **25** (2004), 431–455.

[7] List, J. A. and Gallet, C. A., *The Environmental Kuznets Curve: Does one size fit all?*, Ecological Economics, **31** (1999), 409–423.

[8] Flores, C. A. and Flores-Lagunes, A. and Dimitrios Kapetanakis, *Lessons from Quantile panel estimation of the Environmental Kuznets Curve*, Econometric Reviews, **33** (2014), no. 8, 815–853.

[9] Silverman, Bernard W., *Density Estimation*, Chapman and Hall, London (1986).

[10] Lin, Z., Li, Q. and Sun, Y., *A consistent nonparametric test of parametric regression functional form in fixed effects panel data models*, Journal of Econometrics, **178** (2014), 167–179.

[11] Sun, Y. and Malikov, E., *Estimation and inference in functional-coefficient spatial autoregressive panel data models with fixed effects*, Journal of Econometrics, **203** (2018), no.2, 359–378.

[12] Malikov, E. and Kumbhakar, S. C. and Sun, Y., *Varying Coefficient Panel Data Model in the Presence of Endogenous Selectivity and Fixed Effects*, Journal of Econometrics, **190** (2016), no. 2, 233–251.

[13] Malikov, E. and Kumbhakar, S. C.,*A generalized panel data switching regression model*, Economics Letters, **124** (2014), no. 3, 353-357.

[14] Henderson, D.J. and Parmeter, C. F., *Applied Nonparametric Econometrics*, Cambridge University Press (2015).

[15] Henderson, D. J. and Carroll, R. J. and Li, Q., *Nonparametric estimation and testing of fixed effects panel data models*, Journal of Econometrics, **144** (2008), 257–275.

[16] Henderson, D. J. and Parmeter, C. F., *Panel Data Econometrics*, Cambridge University Press (2015).

[17] Arellano, M., *Panel Data Econometrics*, Oxford University Press (2003).

[18] Baltagi, B., *Econometrics Analysis of Panel Data (2nd edition)*, Wiley, New York (2005).

[19] Hsiao, C., *Analysis of Panel Data (2nd edition)*, Cambridge University Press (2003).

[20] Qian, J. and Wang, L., *Estimating semiparametric panel data models by marginal integration*, Journal of Econometrics, **167** (2012), 483–493.

[21] Li, C. and Liang, Z., *Asymptotics for nonparametric and semiparametric fixed effects panel models*, Journal of Econometrics, **185** (2015) 420–434.

[22] Ullah, A., *Specification analysis of econometric models*, Journal of Quantitative Economics, **1** (1985), 187–209.

[23] Rodriguez-Poo, J. M. and Soberón, A., *Differencing techniques in semi-parametric panel data varying coefficient models with fixed effects: a Monte Carlo study*, Computational Statistics, **30** (2015) , 885–906.

[24] Rodriguez-Poo, J. M. and Soberón, A., *Direct semi-parametric estimation of fixed effects panel data varying coefficient models*, Econometrics Journal, **17** (2014), 107–138.

[25] Rodriguez-Poo, J. M. and Soberón, A., *Nonparametric estimation of fixed effects panel data varying coefficient models*, Journal of Multivariate Analysis, **133** (2015), 95–122.

[26] Millimet, D. L. and List, J. A. and Stengos, T., *The Environmental Kuznets Curve: Real Progress or Misspecified Models?*, Review of Economics and Statistics, **85** (2003), no. 4, 1038–1047.

[27] Bloom, N. and Genakos, C. and Sadun, R. and Van Reenen, J., *Management practices across firms and countries*, NBER Working Paper 17850 (2012).

[28] Bloom, N. and Sadun, R. and Van Reenen, J., *Management as a technology?*, NBER Working Paper 22327 (2017).

[29] Su, L. and Ullah, A., *Profile likelihood estimation of partially linear panel data models with fixed effects*, Economics Letters, **92** (2006), no. 1, 75–81,

[30] Bloom, N. and Van Reenen, J., *Measuring and explaining management practices across firms and countries*, Quarterly Journal of Economics, **122** (2007), no. 4, 1341–1408.

[31] Dewitte, R. and Dumont, and Merlevede, B. and Rayp, G. and Verschelde, M., *Firm-heterogeneous biased technological change: A nonparametric approach under endogeneity*, European Journal of Operational Research, **283** (2020), no. 3, 1172–1182.

[32] Zhang, T. and Wu, W. B., *Inference of time-varying regression models*, Annals of Statistics, **40**(2012), no. 3, 1376–1402.

[33] Cai, Z. and Fang, Y. and Lin, M. and Su, J., *Inferences for a partially varying coefficient model with endogenous regressors*, Journal of Business and Economic Statistics, **37** (2019), no. 1, 158–170.

[34] Neyman, J. and E. Scott, *Consistent estimates based on partially consistent observations*, Econometrica, **16** (1948), 1–32.

[35] Sun, K. and Kumbhakar, S. C., *Semiparametric smooth-coefficient stochastic frontier model*, Economics Letters, **120** (2013), no. 2, 305–309.

[36] Li, Q. and Racine, J. S., *Nonparametric Estimation of Distributions with Categorical and Continuous Data*, Journal of Multivariate Analysis, **86** (2003), 266-292.

[37] Robinson, P. M., *Root-n Consistent Semiparametric Regression*, Econometrica, **56** (1988), 931-954.

[38] Aitchison, J. and Aitken, C. G. G., *Multivariate Binary Discrimination by the Kernel Method*, Biometrika, **63** (1976), 413-420.

[39] Quenouille, M. H., *Notes on Bias in Estimation*, Biometrika, **43** (1956), 353–360.

[40] Tukey, J. W., *Bias and Confidence in Not-Quite Large Samples (Abstract)*, Annals of Mathematical Statistics, **29** (1958), 614.

[41] Newey, W. K., *Estimation of Nonlinear Models in Panel Data*, NSF Proposal, (1988).

[42] Henderson, D.j., Papageorgiou, C. and Parmeter, C. F., *Growth Empirics without Parameters*, Economic Journal, **122** (2012), 125-154.

[43] Henderson, D.j., Papageorgiou, C. and Parmeter, C. F., *Who Benefits from Finiancial Development? New Methods, New Evidence*, European Economic Review, **63** (2013), 47-67.

[44] Lee, Y. Mukherjee, D., *New nonparametric estimation of the marginal effects in fixed-effects panel models: an application on the environmental Kuznets Curve*, Working paper, University of Michigan, (2008).

Appendix

Algorithm for Modified Rodriguez-Poo & Soberon Two-stage Estimator

```
#Modified Rodriguez-Poo and Soberon first-stage estimator
# Gaussian kernel function for continuous regressors
kk < − function(g) { (1/sqrt(2*pi))*exp(-0.5*g2̂) }

firststage < − function(Y,X,Z,X1,Z1,h) {
p < − ncol(X)
q < − ncol(Z)
NT < − nrow(X)
ones < − rep(1,NT)
epsilon < − 0.00001

betahat < − matrix(ncol=2*p*(q+1),nrow=NT)

for (j in 1:NT) {
# Construct NTxq matrix of deviations of Z from z ZZ < − matrix(ncol=q, nrow=NT)
for (jj in 1:q) {
zt < − Z[,jj]-Z[j,jj]
ZZ[,jj] < − zt
}
ZZ1 < − matrix(ncol=q, nrow=NT)
for (jj in 1:q) {
```

```
zt1 <- Z1[,jj]-Z1[j,jj]
ZZ1[,jj] <- zt1
}

# Construct NTxp(q+1) matrix of explanatory variables
RR.aux <- matrix(ncol=p*q,nrow=NT)
for (jj in 1:NT) {
RR.aux[jj,] <- ZZ[jj,]%x%X[jj,]
}
RR <- cbind(X,RR.aux)

RR1.aux <- matrix(ncol=p*q,nrow=NT)
for (jj in 1:NT) {
RR1.aux[jj,] <- ZZ1[jj,]%x%X1[jj,]
}

RR1 <- cbind(X1,RR1.aux)

# Construct NTx1 vector of product kernels
KK.aux <- ones
for (jj in 1:q) {
dz <- (Z[,jj]-Z[j,jj])/h[jj]
KK.aux <- KK.aux*kk(dz)
}
KK1.aux <- ones
for (jj in 1:q) {
dz1 <- (Z1[,jj]-Z1[j,jj])/h[jj]
```

```
KK1.aux < − KK1.aux*kk(dz1)
}
KK < − KK.aux*KK1.aux

RR.com < − cbind(RR,RR1)

# Estimate p(q+1) vector of coefficients and their respective partial derivatives
RK < − crossprod(RR.com,diag(KK))

# Ridging (if necessary)
ridge < − 0
if (tryCatch(as.matrix(solve(RK%*%RR.com+diag(rep(ridge,2*p*(q+1))))),
error=function(e) return(c('error')) )[1]=='error') {
ridge < − ridge + epsilon
}

RKR < − solve(RK%*%RR.com + diag(rep(ridge,2*p*(q+1))))
RKY < − RK%*%Y

betahat[j,] < − RKR%*%RKY
}

return(betahat)

}
```

```
# Modified Rodriguez-Poo and Soberon second-stage estimator

secondstage < − function(Y,X,Z,h) {

NT < − NT-N
T < − T-1
p < − ncol(X)
q < − ncol(Z)

ones < − rep(1,NT)
epsilon < − 0.00001

betahat < − matrix(ncol=p*(q+1),nrow=NT)
for (j in 1:NT) {
Construct NTxq matrix of deviations of Z from z
ZZ < − matrix(ncol=q, nrow=NT)
for (jj in 1:q) {
zt ¡- Z[,jj]-Z[j,jj]
ZZ[,jj] ¡- zt
}

# Construct NTxp(q+1) matrix of explanatory variables
RR.aux < − matrix(ncol=p*q,nrow=NT)
for (jj in 1:NT) {
RR.aux[jj,] < − ZZ[jj,]%x%X[jj,]
}
RR ¡- cbind(X,RR.aux)
```

```
# Construct NTx1 vector of product kernels
KK < − ones
for (jj in 1:q) {
dz < − (Z[,jj]-Z[j,jj])/h[jj]
KK < − KK*kk(dz)
}
#Estimate p(q+1) vector of coefficients and their respective partial derivatives
RK < − crossprod(RR,diag(KK))

# Ridging (if necessary) to find inverse of RSR
ridge < − 0
if (tryCatch(as.matrix(solve(RK%*%RR+diag(rep(ridge,p*(q+1))))),
error=function(e) { return(c('error')) }
)[1]=='error') {
ridge < −ridge + epsilon
}

RKR < − solve(RK%*%RR + diag(rep(ridge,p*(q+1))))
RKY < − RK%*%Y
betahat[j,] < − RKR%*%RKY
}

return(betahat)
}
```

Algorithm for Two-way Fixed Effects Semiparametric Model

```
# Gaussian kernel function for continuous regressors
kk < − function(g) { (1/sqrt(2*pi))*exp(-0.5*g2̂) }

# FE design matrices
# Unrestricted NTxNT design matrix
aux < − diag(rep(1,N))
onesT < − matrix( rep(1,T), ncol=1, nrow=T)
D < − aux%x%onesT

ones.I < − matrix( rep(1,N), ncol=1, nrow=N)
aux.T < − diag(rep(1,T))
P < − ones.I%x%aux.T

# Sun et al. Unrestricted Time Effect estimator

SunetalT.ur < −function(Y,X,Z,h)

p < − ncol(X)
q < − ncol(Z)
ones < − rep(1,NT)
epsilon < − 0.00001

betahat < − matrix(ncol=p*(q+1),nrow=NT)
for (j in 1:NT) {
# Construct NTxq matrix of deviations of Z from z
ZZ < − matrix(ncol=q, nrow=NT)
```

```
for (jj in 1:q) {
zt < − Z[,jj]-Z[j,jj]
ZZ[,jj] < − zt
}

# Construct NTxp(q+1) matrix of explanatory variables
RR.aux < − matrix(ncol=p*q,nrow=NT)
for (jj in 1:NT) {
RR.aux[jj,] < − ZZ[jj,]%x%X[jj,]
}
RR < − cbind(X,RR.aux)

# Construct NTx1 vector of product kernels
KK < − ones
for (jj in 1:q) {
dz < − (Z[,jj]-Z[j,jj])/h[jj]
KK < − KK*kk(dz)
}

# Construct transformation matrix
PK < − matrix(ncol=T, nrow=NT)
for (jj in 1:T) {
PK[,jj] < − P[,jj]*KK
}
PK < − t(PK)
```

```
# Ridging (if necessary) to find inverse of DKD
ridge < − 0
if (tryCatch(as.matrix(solve(PK%*%P+diag(rep(ridge,T)))),
error=function(e) { return(c('error')) } )[1]=='error') {
ridge < − ridge + epsilon
}

PKP < − solve(PK%*%P + diag(rep(ridge,T)))
NN.aux < − PKP%*%PK
NN.aux < − P%*%NN.aux
NN < − diag(rep(1,NT))-NN.aux

NK < − matrix(ncol=NT, nrow=NT)
for (jj in 1:NT) {
NK[,jj] < − NN[,jj]*KK
}
NK < − t(NK)
GG < − NK%*%NN

# Construct NTxNT transformation matrix
DG < − crossprod(D,GG)

ridge < − 0
if (tryCatch(as.matrix(solve(DG%*%D+diag(rep(ridge,N)))),
error=function(e) { return(c('error')) } )[1]=='error') {
ridge < − ridge + epsilon
```

```
}

DGD < − solve(DG%*%D + diag(rep(ridge,N)))
MM.aux < − DGD%*%DG
MM.aux < − D%*%MM.aux
MM < − diag(rep(1,NT))-MM.aux
MG < − crossprod(MM,GG)
SS < − MG%*%MM

#Estimate p(q+1) vector of coefficients and their respective partial derivatives
RS < − t(RR)%*%SS

# Ridging (if necessary) to find inverse of RSR
ridge < − 0
if (tryCatch(as.matrix(solve(RS%*%RR+diag(rep(ridge,p*(q+1))))),
error=function(e) { return(c('error')) }
)[1]=='error') {
ridge < − ridge + epsilon
}

RSR < − solve(RS%*%RR + diag(rep(ridge,p*(q+1))))
RSY < − RS%*%Y

betahat[j,] < − RSR%*%RSY
```

```
    return(betahat)
}
```