2021

A varying coefficient model with two-way fixed effects and different smoothing variables

Taining Wang
Capital University of Economics and Business, taining.wang@cueb.edu.cn

Feng Yao
West Virginia University, Feng.Yao@mail.wvu.edu

A VARYING COEFFICIENT MODEL WITH TWO-WAY FIXED EFFECTS AND DIFFERENT SMOOTHING VARIABLES

Taining Wang[*]
International School of Economics and Management
Capital University of Economics and Business
Beijing, 100070, China

Feng Yao[†]
Department of Economics, West Virginia University, 1601 University Ave., P.O. Box 6025, Morgantown, WV 26505, USA; Voice: +1 304 293 7867

Abstract: We propose a varying coefficient regression model for panel data that controls for both latent heterogeneities in cross-sectional units and unobserved common shocks over time. The model allows different smoothing variables to enter through either a stand-alone function or a coefficient function. Without requiring a normalization of the fixed effects, we propose a two-step estimator. First, we estimate the varying coefficients with the pilot series-based estimators, eliminating fixed effects though differencing. Second, we perform a one-step kernel backfitting to improve the estimation efficiency. We demonstrate through Monte-Carlo simulations that our estimators are computationally efficient and perform well relative to a profile-based kernel estimator.

Keywords: Semiparametric model; varying coefficient model; different smoothing variables; two-way fixed effects; series estimation; kernel backfitting.

JEL Classifcation: C14, C15, C22.

1 Introduction

The linear panel regression model has been a major workhorse in various fields of economics owning to the increasing availability of panel data (Hsiao, 2014; Baltagi, 2013). The panel data model with fixed effects gains wide popularity because it allows a non-zero correlation between latent heterogeneities and covariates, a much needed feature in applied studies. Nonetheless, the linear panel model places restrictive assumptions on the regression functional form, violation of which generally leads to inconsistent estimation and invalid inference. Nonparametric fixed effect models reduce the risk of model misspecification (Henderson et al., 2008; Lin et al., 2014); however, a fully nonparametric model presents a challenge, due to the curse of dimensionality, for empirical applications with a large number of covariates. Semiparametric models provide a viable alternative.[1] Among others, Sun et al. (2009) estimate a varying coefficient panel data model (VCPM) through a profile least-squares estimator (PLS), where the fixed effects are handled through a kernel-based weighting matrix (see also Su and Ullah (2006) and Rodriguez-Poo and Sobern (2014, 2015)). However, most of the work focus on the one-way fixed effects model.

Two-way fixed effects in the panel data model are highly desirable as they accommodates unobserved heterogeneities in *both* cross-sectional (n) and time (T) dimensions that may be correlated with regressors. Estimation of two-way fixed effects VCPM, however, is not trivial. With a short panel frequently encountered in economics where $n \gg T$, one tends to use time-differenced data together with time fixed effects in a partially linear model, which can be theoretically complicated and computationally demanding. Extending Sun et al. (2009), Halder and Malikov (2020) propose a one-step PLS estimator to concentrate out both fixed effects by a local kernel-smoothed two dimensional within-transformation. The PLS estimates coefficient functions of time-varying variables, and performs well in their simulation. We observe that a normalization condition on the fixed effects or a modification is necessary for their estimation even with all variables being time-varying, because their estimator relies on the invertibility of a weighting matrix to concentrate out the two-way fixed effects (see Section 3.2).

Allowing further flexibility in the model can be useful, such as varying coefficient functions associated with either constant regressors (see Wang et al. (2020) and Tian et al. (2021) for their roles in applications), or time-varying regressors, where these coefficients can be functions of the same or different smoothing variables. In the context of production function, for instance, an

[1] See Su and Ullah (2011) and Rodriguez-Poo and Soberon (2017) for a thorough review of recent development in nonparametric and semiparametric panel models.

increasing R&D expenditure may change the marginal productivity of capital, while the air-pollution may negatively impact the marginal productivity of labor. In models without fixed effects, one can apply the two-step marginal integration estimation (Yang et al., 2006; Zhang and Li, 2007; Xue and Yang, 2006), or the smooth-backfitting (Lee et al., 2012,b; Roca-Pardinas and Sperlich, 2010).[2] We are not aware of works on estimating VCPM with different smoothing variables and two-way fixed effects.

In this paper, we propose an estimator for the VCPM and our work exhibits three features. First, we propose a two-step estimator for the VCPM with two-way fixed effects. Second, we estimate coefficient functions of both constant and time-varying variables, without requiring the normalization condition on the fixed effects. Third, we allow the varying coefficients to be potentially functions of different smoothing variables, offering a greater flexibility to model the underlying relationship among covariates. In the first step, we employ a pilot series-based estimator for the varying coefficients. Maintaining the structure of VCPM, we do to need to perform joint smoothing, and we eliminate fixed effects through differencing. In the second step, we perform a one-step kernel backfitting to improve estimation efficiency. Through Monte-Carlo simulations, we show that our proposed estimator works well. Because the fixed effects are eliminated through differencing, we conjecture that our estimator exhibits a simpler bias expression compared to the PLS in Halder and Malikov (2020). This conjecture is supported in simulations, in which we show that our two-step estimator provides a computationally efficient alternative to the PLS estimator. We provide the model and the estimator in Section 2, the simulation study and a comparison with the PLS in Section 3, and conclusions in Section 4.

[2] See also Lee et al. (2019) and Mammen et al. (2009) for interesting work in purely additive model with fixed effects.

2 Model and Estimation

Consider the following VCPM with two-way fixed effects and different smoothing variables

$$Y_{it} = \alpha_i + \mu_t + \sum_{j=1}^{d} \left(m_j(Z_{jit}) + X_{jit}^{\top} \beta_j(Z_{jit}) \right) + u_{it}, \quad i = 1, ..., n, \ t = 1, ..., T, \tag{1}$$

where n and T are the total number of cross-sectional and time units, respectively. Y_{it} is the dependent variable and $Z_{jit} \in \Re^{q_j}$ are the smoothing variables for $j = 1, ..., d$, which enter the coefficient function $m_j(\cdot)$ of the constant and also enter coefficient functions $\beta_j(Z_{jit}) = [\beta_{j1}(Z_{jit}), ..., \beta_{jp_j}(Z_{jit})]^{\top}$ correspondingly of $X_{jit} = [X_{jit,1}, ..., X_{jit,p_j}]^{\top} \in \Re^{p_j}$, with $p_1 + \cdots + p_d = p$. In this paper, we consider univariate Z_{jit} ($q_j = 1$) due to the consideration of the curse of dimensionality, and focus on the case with $d = 2$ for ease of illustration. The unknown and smooth functions $m_j(\cdot)$ and $\beta_{js}(\cdot)$ for $s = 1, ..., p_j$ are of interest, since they capture the stand-alone effect of Z_{jit} and marginal effects of X_{jit} through Z_{jit}. α_i and μ_t are fixed effects that captures latent heterogeneities in cross-sectional units and unobserved common factors over time, respectively. Denote $Z_{it} = [Z_{1it}, Z_{2it}]^{\top} \in \Re^2$, $X_{it} = [X_{1it}^{\top}, X_{2it}^{\top}]^{\top} \in \Re^p$, and $W_{it} = [Z_{it}^{\top}, X_{it}^{\top}]^{\top} \in \Re^{2+p}$. We allow arbitrary correlation between fixed effects and covariates, so $E(\alpha_i|W_{it}, \mu_t) \neq 0$ and $E(\mu_t|W_{it}, \alpha_i) \neq 0$. Finally, the idiosyncratic error $u_{it} \sim iid(0, \sigma_u^2)$ is independent of W_{it}, α_i, and μ_t. We follow a conventional practice in Li (2000) to impose $m_j(0) = 0$ for $j = 1, 2$ to identify the stand-alone additive coefficients.[3]

The flexible structure in (1) relates to many popular non/semiparametric panel models in the literature. If $\mu_t = 0$ and $\alpha_i \neq 0$ (one-way fixed effects), (1) becomes the nonparametric models (Henderson et al., 2008; Gao and Li, 2013; Lin et al., 2014; Lee et al., 2019) when $d = 1$ and $p = 0$; the varying coefficient model with the same smoothing variables (Rodriguez-Poo and Sobern, 2014, 2015) when $d = 1$, $p = p_1$, and $m_j(\cdot) = 0$.[4] If $\mu_t \neq 0$ and $\alpha_i \neq 0$ (two-way fixed effects), (1) reduces to the nonparametric model (Lee

[3]Clearly, the separable identification of α_i and μ_t requires further restrictions. We are not concerned with the latter since the estimator below removes both fixed effects.

[4]See also Ai et al. (2014) for the partially linear additive model with fixed effects.

et al., 2019) when $d = 1$ and $p = 0$; the additive model (Mammen et al., 2009) when $d > 1$ and $p = 0$; and the varying coefficient model with the same smoothing variables (Halder and Malikov, 2020) when $d = 1$, $p = p_1$, and $m_j(\cdot) = 0$.

Suppose $m_j(\cdot)$ and $\beta_{js}(\cdot)$ were known, we can eliminate the fixed effects through a multi-differencing step

$$\ddot{\Delta} Y_{it} = \sum_{j=1}^{2} \left(\ddot{\Delta} m_j(Z_{jit}) + \ddot{\Delta} X_{jit}^\top \ddot{\Delta} \beta_j(Z_{jit}) \right) + \ddot{\Delta} u_{it}, \quad i = 2, ..., n, \; t = 2, ..., T, \tag{2}$$

where $\ddot{\Delta} Y_{it} = Y_{it} - Y_{i(t-1)} - Y_{1t} + Y_{1(t-1)}$ with a total sample size of $N = nT - n - T + 1$, and we assume $E(\ddot{\Delta} u_{it} | W_{it}, W_{i(t-1)}, W_{1t}, W_{1(t-1)}) = 0$. However, the transformation in (2) is infeasible because $m_j(\cdot)$ and $\beta_j(\cdot)$ are unknown. Below, we propose a two-step estimation.

In the first step, we perform a series approximation on nonlinear functions. Let $\mathcal{Z} = [a, b] \subset \Re$ be a common compact support of the random variable Z_j for some constants a and b. Denote $\phi(z_j) = [\phi_1(z_j), ..., \phi_{L_N}(z_j)]^\top$ as a $L_N \times 1$ basis functions evaluated at $z_j \in \mathcal{Z}$. Approximately, $m_j(z_j) \approx \phi(z_j)^\top \lambda_j^m$, for some series coefficients $\lambda_j^m = [\lambda_{j,1}^m, ..., \lambda_{j,Ln}^m]^\top$.[5] Similarly, $\beta_{js}(z_j) \approx \phi(z_j)^\top \lambda_{js}^\beta$, where $\lambda_{js}^\beta = [\lambda_{js,1}^\beta, ..., \lambda_{js,L_N}^\beta]^\top$ are unknown series coefficients. With $\lambda_j^\beta = [\lambda_{j1}^{\beta\top}, ..., \lambda_{jp_j}^{\beta\top}]^\top$ and $\lambda_j = [\lambda_j^{m\top}, \lambda_j^{\beta\top}]^\top \in \Re^{(1+p_j)L_N}$, we obtain $m_j(Z_{jit}) + X_{jit}^\top \beta_j(Z_{jit}) \approx \psi_{jit}^\top \lambda_j$, where $\psi_{jit} = [\phi(Z_{jit})^\top, X_{jit}^\top \otimes \phi(Z_{jit})^\top]^\top$ and $\otimes$ denotes the Kronecker product. Thus, (2) becomes

$$\ddot{\Delta} Y_{it} \approx \ddot{\Delta} \psi_{it}^\top \lambda + \ddot{\Delta} u_{it}, \tag{3}$$

where $\ddot{\Delta} \psi_{it} = [\ddot{\Delta} \psi_{1it}^\top, \ddot{\Delta} \psi_{2it}^\top]^\top \in \Re^{(2+p)L_N}$, $\ddot{\Delta} \psi_{jit} = [\ddot{\Delta} \phi(Z_{jit})^\top, \ddot{\Delta}(X_{jit}^\top \otimes \phi(Z_{jit})^\top)]^\top$, and $\lambda = [\lambda_1^\top, \lambda_2^\top]^\top$. Let $\ddot{\Delta} Y_i = [\ddot{\Delta} Y_{i2}, ..., \ddot{\Delta} Y_{iT}]^\top$ and $\ddot{\Delta} u_i = [\ddot{\Delta} u_{i2}, ..., \ddot{\Delta} u_{iT}]^\top$ be $(T \times 1)$ vector, and $\ddot{\Delta} \psi_i = [\ddot{\Delta} \psi_{i2}, ..., \ddot{\Delta} \psi_{iT}]$ be a $(2 + p)L_N \times (T - 1)$ matrix. We rewrite (3) in matrix form as

$$\ddot{\Delta} Y = \ddot{\Delta} \Psi \lambda + \ddot{\Delta} u, \tag{4}$$

where $\ddot{\Delta} Y = [\ddot{\Delta} Y_2^\top, ..., \ddot{\Delta} Y_n^\top]^\top$, $\ddot{\Delta} u = [\ddot{\Delta} u_2^\top, ..., \ddot{\Delta} u_n^\top]^\top$, each being a $N \times 1$ vector, and $\ddot{\Delta} \Psi = [\ddot{\Delta} \psi_2, ..., \ddot{\Delta} \psi_n]^\top$ is a $N \times (2 + p)L_N$ matrix. Our series

[5] For identification, we impose $\phi_l(z_j = 0) = 0$ with a proper choice of basis functions.

coefficient estimator is obtained via the least squares from

$$\hat{\lambda} = \underset{\lambda \in \Re^{(2+p)L_N}}{\operatorname{argmin}} [\ddot{\Delta} Y - \ddot{\Delta} \Psi \lambda]^\top [\ddot{\Delta} Y - \ddot{\Delta} \Psi \lambda], \tag{5}$$

and we obtain the corresponding coefficient function estimates as $\hat{m}_j(z_j) = \phi(z_j)^\top \hat{\lambda}_j^m$ and $\hat{\beta}_{js}(z_j) = \phi(z_j)^\top \hat{\lambda}_{js}^\beta$, respectively.

Remarks: First, different from the initial step in marginal integration, the series estimate does not require a full-dimensional nonparametric smoothing. Thus, it is computationally efficient. Second, the series estimator can easily impose the differenced model structure in (2), which removes the two-way fixed effects, and maintains the varying coefficient structure with different smoothing variables. Third, the differenced estimate calls for determination of the first individual unit at $i = 1$ as a reference point (e.g., Y_{1t}). Though there are situations where the first unit can be assigned from the context, determining the first unit is mostly arbitrary in nature. We reduce the arbitrariness of assignments by relying only on the first unit to define the estimator. With data being independently and identically distributed across the individual units, the impact of the choice of the first unit on the property of the estimator might be negligible. However, the estimate value will vary with a different choice of the first unit. We show through simulations below that the choice of the first unit, though arbitrary, does not seem to disturb the performance of the estimator qualitatively.

In the second step, using the pilot series-based estimates, we perform a one-step backfitting for $m_j(\cdot)$ and $\beta_{js}(\cdot)$ by local linear estimation to improve efficiency. Let $\chi_{jit} = [1, X_{jit}^\top]^\top$, $g_j(z_j) = [m_j(z_j), \beta_j(z_j)^\top]^\top$, and define $\ddot{\Delta} Y_{it}^{(j)} = \ddot{\Delta} Y_{it} - \sum_{j'=1, j' \neq j}^{2} \ddot{\Delta}(\chi_{j'it}^\top g_{j'}(Z_{j'it})) + \chi_{ji(t-1)}^\top g_j(Z_{ji(t-1)}) + \chi_{j1t}^\top g_j(Z_{j1t}) - \chi_{j1(t-1)}^\top g_j(Z_{j1(t-1)})$,

$$\ddot{\Delta} Y_{it}^{(j)} = \chi_{jit}^\top g_j(Z_{jit}) + \ddot{\Delta} u_{it}, \tag{6}$$

where the infeasible dependent variable is the difference of the left-hand in (2) and the right-hand regression functions except these depending on Z_{jit}. In practice, we use $\ddot{\Delta} \hat{Y}_{it}^{(j)}$, where the unknown $g_{j'}(\cdot)$ and $g_j(\cdot)$ are replaced by their series estimates $\hat{g}_{j'}(\cdot)$ and $\hat{g}_j(\cdot)$ in the first step. Let $k(v)$ be a univariate kernel function, and define the $N \times 1$ matrix $\ddot{\Delta} \hat{Y}^{(j)} = \{\ddot{\Delta} \hat{Y}_{it}^{(j)}\}_{i=2,t=2}^{n,T}$, the $N \times N$ diagonal kernel weighting matrix $K(z_j) = diag\{k\left(\frac{Z_{jit} - z_j}{h}\right)\}_{i=2,t=2}^{n,T}$, and the $N \times (2+2p_j)$ matrix $R(z_j) = \{R_{it}(z_j)^\top\}_{i=2,t=2}^{n,T}$ with $R_{it}(z_j) = [\chi_{jit}^\top, \chi_{jit}^\top (Z_{jit} - z_j)]^\top$. In (6), we estimate the functions $g_j(z_j)$ by $\check{g}_j(z_j) \equiv \check{c}_{0j} \in \Re^{1+p_j}$, and the first-order derivatives $g_j'(z_j) = [m_j'(z_j), \beta_j'(z_j)^\top]^\top$ by $\check{g}_j'(z_j) \equiv \check{c}_{1j} \in \Re^{1+p_j}$, where $\check{c}_j = [\check{c}_{0j}^\top, \check{c}_{1j}^\top]^\top \equiv \check{c}_j(z_j)$ are local linear estimates from

$$\check{c}_j(z_j) = \underset{c_j(z_j) \in \Re^{2+2p_j}}{\operatorname{argmin}} \left[\ddot{\Delta}\hat{Y}^{(j)} - R(z_j)c_j(z_j)\right]^\top K(z_j) \left[\ddot{\Delta}\hat{Y}^{(j)} - R(z_j)c_j(z_j)\right]. \tag{7}$$

We denote the estimator in (7) as DKB for differencing kernel backfitting estimator, which differs from the PLS estimator by Halder and Malikov (2020) in three aspects. First, the PLS focuses on the case of (1) with $d = 1$ and $p = p_1$, whereas we allow for a more general case of $d > 1$ and $p > p_1$ with different smoothing variables. Second, the PLS is a one-step estimator that estimates only $\beta_j(\cdot)$ in $g_j(\cdot)$ by *concentrating out* the fixed effects through a profile approach. The profile approach as in Sun et al. (2009) results in a bias term for the estimator that can be involved, and we conjecture a similar expression to appear in Halder and Malikov (2020). We estimate $g_j(\cdot)$ by (7) in the *absence* of fixed effects, because they are differenced out using the series estimator in the first step. Different methods of treating the fixed effects might lead to estimators with different finite sample performance, which we explore in the simulations. Third, the PLS estimation in the two-way effects requires the fixed effects to satisfy a normalization condition (discussed in Section 3.2), or a modification on estimates. In contrast, we do not need such a condition to estimate both $m_j(\cdot)$ and $\beta_{js}(\cdot)$.

3 Monte-Carlo Studies

In Section 3.1, we evaluate the finite sample performance of DKB in (7), and in Section 3.2, we perform a comparison study for DKB and PLS by Halder and Malikov (2020).

3.1 Performance of the proposed estimator

We consider the following data generating process (DGP)

$$Y_{it} = \alpha_i + \mu_t + \sum_{j=1}^{2} \Big(m_j(Z_{jit}) + X_{jit}\beta_j(Z_{jit}) \Big) + u_{it}, \tag{8}$$

where $d = 2$ and $p_1 = p_2 = 1$, so $X_{it} = [X_{1it}, X_{2it}]^\top$, $Z_{it} = [Z_{1it}, Z_{2it}]^\top$, and $\beta_{js}(\cdot) = \beta_j(\cdot)$. We generate $W_{it} = [Z_{it}^\top, X_{it}^\top]^\top$ from a multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$, where μ is a 4×1 vector with each element being 2, and Σ is a covariance matrix with its ij^{th} element given by $\Sigma_{ij} = \rho^{|i-j|}$ for $i, j = 1, ..., 4$. We set $\rho = 0.5$ to allow covariates to be correlated, and generate Z_{jit} with values in $[0, 4]$. The fixed effects $\alpha_i = \frac{1}{T}\sum_{t=1}^{T} c_{01}(\sum_{j=1}^{4} W_{jit}) + \xi_{1i}$ and $\mu_t = \frac{1}{n}\sum_{i=1}^{n} c_{02}(\sum_{j=1}^{4} W_{jit}) + \xi_{2t}$, where $\xi_{1i} \sim \mathcal{N}(0, 1.25^2)$, $\xi_{2t} \sim \mathcal{N}(0, 0.25^2)$, and we set $c_{0k} = 1$ to reflect the fixed effects correlated with regressors for $k = 1, 2$. We draw $u_{it} \sim \mathcal{N}(0, 1)$ independently. We consider low-frequency functions in DGP_1 with $m_1(v) = v$, $m_2(v) = v^2$, $\beta_1(v) = 2.5 + e^{v-1} - v^2$, and $\beta_2(v) = 2e^{-2v} + \frac{1}{2}v^3$; high-frequency functions in DGP_2 with $m_1(v) = v^2$, $m_2(v) = v - v^3$, $\beta_1(v) = -0.5 + \cos(3v\pi/4)$, and $\beta_2(v) = 1 + e^{v-1}\sin(v\pi/2)$.

In the first step, we employ the B-spline estimator with a total of L_N basis functions ($L_N = J_N + m + 1$), where m is the polynomial order of the basis function, and J_N is the number of interior knots placed on the support of Z_j except at boundary. Following a conventional practice in the literature, we implement a cubic B-spline function ($m = 3$) and place the l^{th} interior knot on the $(l/J_N + 1)^{th}$ percentile of Z_j for $l = 1, ..., J_N$. We follow Ai et al. (2014) to select J_N via the cross-validation criterion (CV). In the second step, we use the Gaussian kernel function in (7) with a bandwidth $h_j = C_{h_j}\hat{\sigma}_j N^{-1/5}$, where C_{h_j} is a scaling factor selected by CV, and $\hat{\sigma}_j$ is the standard deviation of Z_{jit}. We choose $n = (100, 200, 400)$, fix $T = (5, 10)$, and perform $2,000$ repetitions. To evaluate the performance of the estimator at 50 fixed grid points in $[0, 4]$, we report the root average MSE (RAMSE), the average absolute bias (ABIAS), and the average standard deviation (ASD).

We provide results for DKB in Table 1 for $T = 5$ in the upper panel and for $T = 10$ in the lower panel. To save space, we report the estimation results for $m_1(Z_1)$, $\beta_2(Z_2)$, and the sum of all functions $\sum_{j=1}^{2}(m_j(\cdot) + \beta_j(\cdot))$

(Total: DKB). We observe that $m_1(Z_1)$ is easier to estimate than $\beta_2(Z_2)$ across both DGPs, judged by the smaller RAMSE of $\check{m}_1(\cdot)$. The performance of DKB clearly improves with a larger sample size across experiment designs, with either a larger n and fixed T, or a larger T with fixed n. To illustrate the impact of different choices of the first unit on the performance of DKB, we report the mean and standard deviation (SD) of RAMSE across n choices of the first unit for $T = 5$ in DGP_1. When $n = 100$, the means of RAMSE for ($\check{m}_1$, $\check{\beta}_2$, Total: DKB) are $(0.3060, 0.4634, 0.8788)$, and the SD $(0.0108, 0.0074, 0.0193)$. When n=400, the mean are $(0.1935, 0.2729, 0.5277)$, and the SD $(0.0021, 0.0006, 0.0008)$. Compared with results in Table 1, DKB's performance does not change qualitatively, suggesting a negligible role for the choice of the first unit.

Figure 1 illustrates vividly the DKB estimates in DGP_{1-2} for $m_1(\cdot)$ (left) and $\beta_2(\cdot)$ (middle), where the true function (black solid line) is plotted together with DKB estimates (blue dash line) using $(n, T) = (400, 10)$. Consistent with the numerical results, the DKB estimate captures the shape of each function well, where the difference is only noticeable for areas with a sharp curvature in $\beta_2(\cdot)$ under DGP_2. In Table 1, we further estimate the sum of all functions by DKB with one-way fixed effects (Total: FD), eliminating only α_i through first-differencing. As expected, the ignorance of the time effects in (Total: FD) significantly deteriorates its performance, judged by the large magnitudes of all measures, ABIAS in particular, which continue to be relatively large as sample size increases. Overall, the simulations indicate an encouraging performance of the DKB, highlighting the importance of controlling two-way fixed effects in panel data.[6]

3.2 A comparison study

Halder and Malikov (2020) consider a special case of VCPM in (1) with $d = 1$, $p = p_1$, and $m(\cdot) = 0$ as

$$Y_{it} = \alpha_i + \mu_t + X_{it}^\top \beta(Z_{it}) + u_{it}. \tag{9}$$

[6]We also perform simulation studies with a higher dimension with $d = 3$, $p_1 = p_2 = p_3 = 1$ or $p_1 = 1$, $p_2 = 2$, and $p_3 = 3$. The results remain qualitatively similar to those reported here. The full simulation results are available upon request.

The model can be appropriate when the same Z appear in all coefficient functions of $X \in \Re^p$, and Z plays no role through $m(\cdot)$. Halder and Malikov (2020) implement the PLS in (9) to concentrate out both fixed effects, and estimate $\beta(\cdot)$ without normalizing fixed effects (i.e., $\sum_{i=1}^n \alpha_i = 0$). We compare of performance of DKB and PLS under (9) below.

Before proceeding, we show that the PLS in Halder and Malikov (2020) does require a normalization of fixed effects or a modification of the estimator. Define $\alpha = [\alpha_1, ..., \alpha_n]^\top$, $\mu = [\mu_1, ..., \mu_T]^\top$, and let $D = I_n \otimes \iota_T$ and $Q = \iota_n \otimes I_T$, where I_e and ι_e are an $e \times e$ identify matrix and an $e \times 1$ vector of ones, respectively, for any positive integer e. (9) in the matrix form is

$$Y = D\alpha + Q\mu + B(X, \beta(Z)) + u, \tag{10}$$

where $Y = \{Y_{it}\}_{i=1}^{n}{}_{t=1}^{T}$, $u = \{u_{it}\}_{i=1}^{n}{}_{t=1}^{T}$ are $nT \times 1$, and $B(X, \beta(Z))$ is a $nT \times 1$ vector with its it^{th} element being $X_{it}^\top \beta(Z_{it})$. Halder and Malikov (2020) concentrate out μ and α to obtain the PLS estimator $\tilde{\beta}(z) \equiv \tilde{c}_0(z) \in \Re^p$, where for $\tilde{c}_1(z) \equiv \tilde{\beta}'(z) \in \Re^p$, $\tilde{c}(z) = [\tilde{c}_0(z)^\top, \tilde{c}_1(z)^\top]^\top$ are obtained from

$$\tilde{c}(z)^\top = \underset{c(z) \in \Re^{2p}}{\operatorname{argmin}} \; [Y - R(z)c(z)]^\top \Sigma(z) [Y - R(z)c(z)]. \tag{11}$$

Here, we assume $q = 1$. $R(z) = \{X_{it}^\top, X_{it}^\top (Z_{it} - z)\}_{i=1,t=1}^{n,T}$ is a $nT \times 2p$ matrix because there are no constant functions in (9). The PLS replaces the weighting matrix $K(z)$ in (7) with $\Sigma(z)$ in (11), where $\Sigma(z) = \Omega_Q(z)^\top S_Q(z) \Omega_Q(z)$, with $\Omega_Q(z) = I_{nT} - D(D^\top S_Q(z) D)^{-1} D^\top S_Q(z)$, $S_Q(z) = M_Q(z)^\top K(z) M_Q(z)$, and $M_Q(z) = I_{nT} - Q(Q^\top K(z) Q)^{-1} Q^\top K(z)$. Intuitively, $\Sigma(z)$ wipes out time effects through $M_Q(z)Q\mu = 0$ and individual effects through $\Omega_Q(z) D\alpha = 0$.

The PLS estimator in (11) requires $\Sigma(z)$ to be well defined. However, we argue that this does require additional conditions, because the matrix $G = D^\top S_Q(z) D$ in $\Sigma(z)$ is singular for the following reasons. First, define $k_{it} \equiv k\left(\frac{Z_{it} - z}{h}\right)$ and note that $G = D^\top K(z) M_Q(z) D$. By simplification, we obtain that $G = \{G_{ij}\}_{i=1,j=1}^{n,n}$ is a $n \times n$ square matrix, with the ij^{th} element G_{ij} being $\sum_{t=1}^T \left(k_{it}\left(\frac{\sum_{s'=1,s' \neq i}^n k_{s't}}{\sum_{s=1}^n k_{st}}\right)\right)$ for $i = j$, and $-\sum_{t=1}^T \left(k_{it}\left(\frac{k_{jt}}{\sum_{s=1}^n k_{st}}\right)\right)$ for $i \neq j$, and $i, j = 1, ..., n$. Thus, $\sum_{j=1}^n G_{ij} = 0$ for all i, indicating the column linear dependence of G.

With a normalization of fixed effects, PLS estimator is well defined. Thus, we compare the DKB and PLS by imposing $\sum_{i=1}^{n} \alpha_i = 0$ in a simulation study. We replace D in (10) by $D_{-1} = [-\iota_{n-1}, I_{n-1}]^\top \otimes \iota_T$, α by $\alpha_{-1} = [\alpha_2, ..., \alpha_n]^\top$, and α_1 by $-\sum_{i=2}^{n} \alpha_i$. We consider the following DGP

$$Y_{it} = \alpha_i + \mu_t + X_{1it}\beta_1(Z_{it}) + X_{2it}\beta_2(Z_{it}) + u_{it}, \tag{12}$$

where all data and DGP_{1-2} are generated the same as before. Our DKB estimator still requires a two-step estimation to eliminate the fixed effects with pilot estimates at different time and individual unit. We select the interior knots J_N based on the same rule before. For a fair comparison, we use the same rule-of-thumb bandwidth $h = C_h\hat{\sigma}(nT)^{-\frac{1}{5}}$ for both DKB and PLS, where $C_h = 1$ and $\hat{\sigma}$ is the standard deviation of Z_{it}. We notice that the PLS estimator can be computationally expensive due to the need to invert matrices with large dimensions. We follow Halder and Malikov (2020) to reduce the sample size to $n = (50, 100, 200)$, fix $T = 3$ and $T = 5$, and perform 500 repetitions.

We report the simulation results in Table 2. As expected, both estimators perform better with a large sample size, indicating the consistency. The DKB outperforms the PLS uniformly across the experiment designs considered, partially due to the simpler bias expression in DKB that we conjecture. The performance discrepancy also seems to persist when either n or T doubles. The third panel in Figure 1 gives the boxplots, illustrating the distribution of RAMSE for estimating the sum of $\beta_1(Z_1)+\beta_2(Z_2)$ in DGP_1 (upper panel) and DGP_2 (lower panel). The DKB (blue box) clearly exhibits an overall better performance over the PLS (yellow box) across both DGPs across different sample sizes.

We further comment on the difference in computational time. Consider the estimation in DGP_1 at 50 grid points as an example. Running R with an Intel Xeon E3-1505M V5 4-core CPU and 64GB RAM for 500 repetitions, for a small sample size of $(n, T) = (50, 3)$, each repetition takes only 0.007s (seconds) on average for DKB compared to 0.365s for PLS. For a relatively large sample size of $(n, T) = (100, 3)$, it takes 0.008s for DKB but 2.631s for PLS. The difference becomes significantly larger when $(n, T) = (200, 5)$, as it takes 0.011s for DKB but 97.246s for PLS, or about 1.65m (minutes).

In fact, 1.65m is long enough for DKB to handle with a huge sample size of $(n, T) = (38,000, 200)$. Overall, the DKB appears attractive in terms of numerical performance and computational efficiency. Thus, the DKB serves as a viable alternative for VCPM with two-way fixed effects and different smoothing variables.

4 Conclusion

Our proposed VCPM controls for two-way fixed effects, allows coefficient functions of both time-varying and constant regressors, and improves the flexibility of coefficient functions by allowing potentially different smoothing variables. Our estimator involves a combined use of series and kernel estimation to eliminate fixed effects through differencing and improve efficiency through one-step backfitting, without requiring a normalization of fixed effects. Monte-Carlo simulations illustrate encouraging performance of the proposed estimator relative to the profile kernel estimator. Our estimator can serve as a practical alternative in applied studies, due the ease and flexibility of implementation and computational efficiency with a large panel data set.

References

Ai, C., You, J., Zhou, Y., 2014. Estimation of fixed effects panel data partially linear additive regression models. The Econometrics Journal 17 (1), 83–106.

Baltagi, B. H., 2013. Econometric Analysis of Panel Data, 2nd Edition. Wiley, New York.

Gao, Y., Li, K., 2013. Nonparametric estimation of fixed effects panel data models. Journal of Nonparametric Statistics 25 (3), 679–693.

Halder, S. C., Malikov, E., 2020. Smoothed LSDV estimation of functional-coefficient panel data models with two-way fixed effects. Economics Letters 192, 109239.

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

Hsiao, C., 2014. Analysis of Panel Data, 3rd Edition., Econometric Society Monographs, New York.

Lee, Y., Mukherjee, D., Ullah, A., 2019. Nonparametric estimation of the marginal effect in fixed-effect panel data models. Journal of Multivariate Analysis 171, 53–67.

Lee, Y. K., Mammen, E., Park, B. U., 2012. Flexible generalized varying coefficient regression models. Annals of Statistics 40, 1906–1933.

Lee, Y. K., Mammen, E., Park, B. U., 2012b. Projection-type estimation for varying coefficient regression models. Bernoulli 18 (1), 177–205.

Li, Q., 2000. Efficient estimation of additive partially linear models. International Economic Review 41 (4), 1073–1092.

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

Mammen, E., Støve, B., Tjøstheim, D., 2009. Nonparametric additive models for panels of time series. Econometric Theory 25 (2), 442–481.

Roca-Pardinas, J., Sperlich, S., 2010. Feasible estimation in generalized structured models. Statistics and Computing 20, 367–379.

Rodriguez-Poo, J. M., Sobern, A., 2014. Direct semi-parametric estimation of fixed effects panel data varying coefficient models. The Econometrics Journal 17 (1), 107–138.

Rodriguez-Poo, J. M., Sobern, A., 2015. Nonparametric estimation of fixed effects panel data varying coefficient models. Journal of Multivariate Analysis 133, 95–122.

Rodriguez-Poo, J. M., Soberon, A., 2017. Nonparametric and semiparametric panel data models: Recent developments. Journal of Economic Surveys 31 (4), 923–960.

Su, L., Ullah, A., 2006. Profile likelihood estimation of partially linear panel data models with fixed effects. Economics Letters 92, 75–81.

Su, L., Ullah, A., 2011. Nonparametric and semiparametric panel econometric models: estimation and testing. Handbook of empirical economics and finance, 455–497.

Sun, Y., Carroll, R. J., Li, D., 2009. Semiparametric estimation of fixed-effects panel data varying coefficient models. In: Nonparametric econometric methods. Emerald Group Publishing Limited, pp. 101–129.

Tian, J., Wang, T., Hall, J., 2021. The effect of exports on labour share: A semiparametric approach using Chinese manufacturing panel data. The World Economy, 1-39.

Wang, T., Tian, J., Yao, F., 2020. Does high leverage ratio influence Chinese firm performance? a semiparametric stochastic frontier approach with zero-inefficiency. Empirical Economics 61, 587–636.

Xue, L., Yang, L., 2006. Estimation of semiparametric additive coefficient model. Journal of Statistical Planning and Inference 136, 2506–2534.

Yang, L., Park, B. U., Xue, L., Härdle, W. K., 2006. Estimation and testing for varying coefficients in additive models with marginal integration. Journal of the American Statistical Association 101 (475), 1212–1227.

Zhang, R., Li, G., 2007. Averaged estimation of functional-coefficient regression models with different smoothing variables. Statistics & Probability Letters 77, 455–461.

Figure 1: Simulated Function Estimation and Comparison in DGP_{1-2}

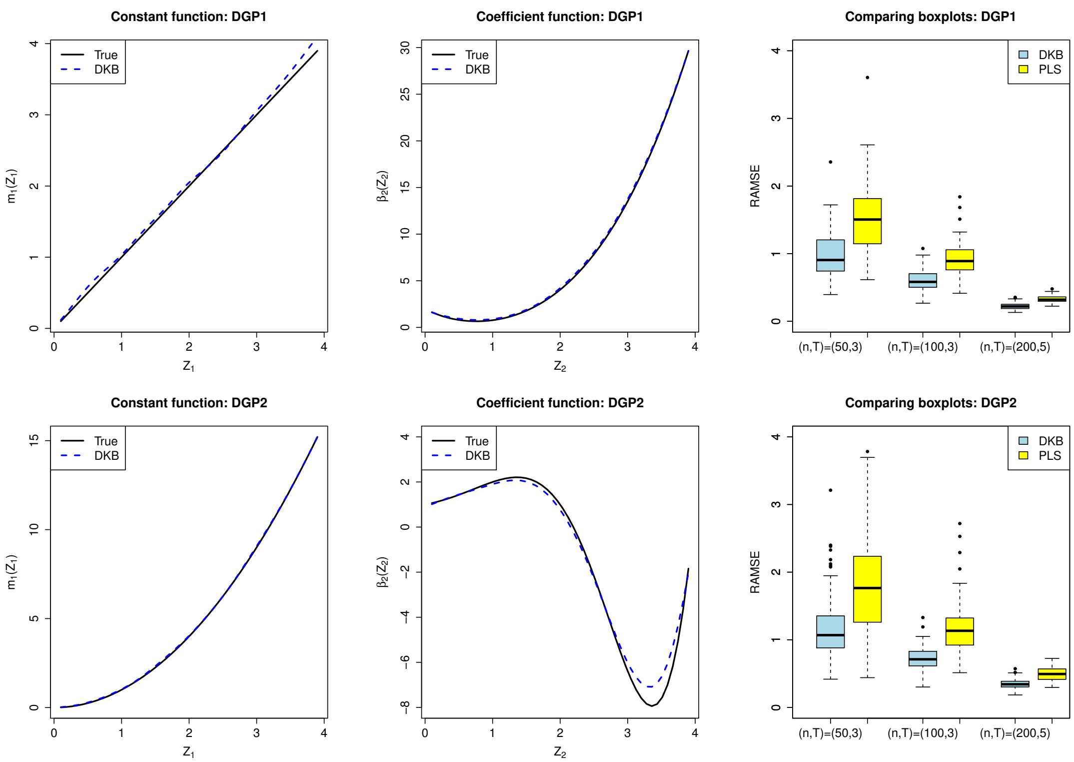

Table 1: Simulation Results with $d = 2$ and $p_1 = p_2 = 1$

		$n = 100$		200		400	
$T = 5$		DGP_1	DGP_2	DGP_1	DGP_2	DGP_1	DGP_2
$\check{m}_1$	RAMSE	0.2861	0.3345	0.1919	0.2887	0.1483	0.1805
	ABIAS	0.2302	0.2776	0.1495	0.2314	0.1141	0.1435
	ASD	0.2827	0.2526	0.1892	0.1966	0.1464	0.1396
$\check{\beta}_2$	RAMSE	0.4169	0.6852	0.3178	0.5781	0.2493	0.4721
	ABIAS	0.3757	0.4921	0.2866	0.4031	0.2270	0.3232
	ASD	0.1106	0.1372	0.0783	0.0919	0.0532	0.0630
Total: DKB	RAMSE	0.9134	0.8661	0.6757	0.6360	0.5348	0.5362
	ABIAS	0.7657	0.6860	0.5716	0.5099	0.4534	0.4332
	ASD	0.4079	0.3775	0.2932	0.2689	0.2135	0.2082
Total: FD	RAMSE	3.0279	2.9080	2.8039	2.4193	1.9544	1.8134
	ABIAS	2.2211	2.0449	2.0904	2.0663	1.8642	1.7326
	ASD	2.4811	2.6038	1.6256	1.4465	0.6601	0.6516
$T = 10$							
$\check{m}_1$	RAMSE	0.1935	0.2597	0.1654	0.2106	0.1091	0.1383
	ABIAS	0.1546	0.2101	0.1317	0.1696	0.0861	0.1134
	ASD	0.1913	0.2005	0.1646	0.1567	0.1088	0.1098
$\check{\beta}_2$	RAMSE	0.3072	0.5522	0.2387	0.4565	0.1833	0.3796
	ABIAS	0.2781	0.3857	0.2173	0.3102	0.1674	0.2537
	ASD	0.0786	0.0878	0.0532	0.0633	0.0385	0.0457
Total: DKB	RAMSE	0.6494	0.6703	0.5241	0.5145	0.3882	0.3914
	ABIAS	0.5433	0.5363	0.4419	0.4137	0.3230	0.3156
	ASD	0.2948	0.3198	0.2485	0.2347	0.1749	0.1609
Total: FD	RAMSE	2.6652	2.3882	2.4033	2.1154	1.8896	1.7887
	ABIAS	1.9942	1.8532	1.7915	1.7473	1.5565	1.5177
	ASD	1.8526	2.1221	1.0113	0.9054	0.5757	0.5233

Table 2: Simulation Results for Comparison between DKB and PLS

DGP_1		$n = 100$		200		400	
$T = 3$		DKB	PLS	DKB	PLS	DKB	PLS
β_1	RAMSE	0.3939	0.4930	0.3037	0.3700	0.2334	0.2872
	ABIAS	0.3001	0.3678	0.2352	0.2755	0.1758	0.2149
	ASD	0.2304	0.2863	0.1508	0.1868	0.1038	0.1303
β_2	RAMSE	0.7109	0.8887	0.5278	0.6521	0.4032	0.5097
	ABIAS	0.6199	0.7779	0.4701	0.5805	0.3627	0.4584
	ASD	0.2557	0.3248	0.1641	0.2061	0.1058	0.1349
$T = 6$							
β_1	RAMSE	0.3298	0.3639	0.2494	0.2964	0.1886	0.2210
	ABIAS	0.2456	0.2676	0.1841	0.2144	0.1413	0.1640
	ASD	0.1651	0.1690	0.1089	0.1111	0.0738	0.0757
β_2	RAMSE	0.5462	0.6415	0.4172	0.4996	0.3239	0.3887
	ABIAS	0.4866	0.5729	0.3741	0.4495	0.2922	0.3535
	ASD	0.1678	0.1792	0.1153	0.1194	0.0744	0.0769
DGP_2							
$T = 3$							
β_1	RAMSE	0.4137	0.6209	0.3135	0.4330	0.2338	0.3036
	ABIAS	0.3387	0.4620	0.2573	0.3338	0.1946	0.2452
	ASD	0.2601	0.3961	0.1831	0.2495	0.1228	0.1621
β_2	RAMSE	0.9742	1.2213	0.8399	1.0039	0.6840	0.8304
	ABIAS	0.7050	0.8752	0.6081	0.7226	0.4875	0.5956
	ASD	0.3312	0.4046	0.2077	0.2821	0.1413	0.1788
$T = 6$							
β_1	RAMSE	0.3164	0.3744	0.2236	0.2846	0.1793	0.2131
	ABIAS	0.2580	0.2992	0.1859	0.2344	0.1512	0.1803
	ASD	0.1864	0.2114	0.1244	0.1422	0.0870	0.0916
β_2	RAMSE	0.8145	0.9769	0.6800	0.7986	0.5722	0.6806
	ABIAS	0.5968	0.7082	0.4866	0.5736	0.3980	0.4756
	ASD	0.2272	0.2638	0.1345	0.1539	0.0923	0.1056