Series Estimation of Functional-Coefficient Partially Linear Regression Model

Kien C. Tran[*]
Department of Economics
University of Lethbridge
4401 University Drive W
Lethbridge, Alberta
T1K 3M4 CANADA

and

Efthymios G. Tsionas
Department of Economics
Lancaster University Management School
LA1 4YX U.K.

Abstract

This paper develops an alternative and complement estimation procedure for functional coefficient partially linear regression (FCPLR) model based on series method. We derive the convergence rates and asymptotic normality of the proposed estimator. We examine its finite sample performance and compare it with the two-step local linear estimator via a small scale Monte Carlo simulation.

Keywords: Functional-coefficient, Series approximation, Convergence rate, Asymptotic normality.

1. Introduction

Functional coefficient models have increasingly become more popular in applied research for various disciplines including among others, economics and finance due to their flexibility and interpretability. For an excellent overview on the methodological and theoretical development on the functional coefficient models, see for example Fan and Zhang (2008).

   Despite their popularity, most of the work thus far has focused mainly on assumption that the coefficient functions have the same smoothing variables, albeit various degree of smoothness is allowed. In practice, this can be a restrictive assumption. To relax this assumption, recently, Wong et al. (2008) extend the functional coefficient models to allow for the functional coefficient functions to depend on different smoothing variables. Their model which they called functional coefficient partially linear regression (FCPLR) model which can be written as:

$$y_i = \theta(u_i) + x_i' \beta(z_i) + e_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $y_i \in \Re$ is the dependent variable, $x_i \in \Re^d$ is a vector of explanatory variables excluding constant term, $u_i \in \Re^p$ and $z_i \in \Re^q$ are vectors of covariates assumed to be exogenous, e_i is a random error assumed to be *i.i.d.* with zero mean and variance σ^2; $\theta(.)$ and $\beta(.)$ are some measurable (unknown) functions. For identification purpose, it is assumed that the elements in u are not contained in x and z. It is clear from (1) that it includes other interesting models as special cases. When $\theta(u) = u'\gamma$, it reduces to the partially linear varying coefficient model, and when $\beta_j(.) = 0, j = 1, \ldots, d$, it becomes the nonparametric regression model. Model (1) also includes the partial linear regression model when $\beta_j(.) = \beta_j$, for all j and additive model when all $x_j = 1$. Finally, if we let $X_i = (1, x_i')'$, $\gamma(u_i, z_i) = (\theta(u_i), \beta(z_i)')'$ and rewrite model (1) as $y_i = X_i'\gamma(u_i, z_i) + e_i$ then model (1) can be viewed as a generalization of the varying coefficient model by allowing for different covariates in different coefficient functions. Thus, model (1) generalizes the varying coefficient models to allow for even greater flexibility and interpretability.

   To estimate the unknown coefficient functions, Wong et al. (2008) propose to use two-step local linear fitting kernel based approach coupled with one-step back-fitting algorithm. They also

derive the asymptotic properties for their estimator. However, albeit local linear estimator is known to possess desirable properties, their estimation procedure involves two-step procedure where in each step, different bandwidths are used and in addition, certain restrictions need to be imposed on the bandwidths between the two-step algorithms. Consequently, this make their estimation procedure more complicated.

Alternatively, one can estimate the unknown functions in model (1) using general series method such as spline or power series, which requires only a single step estimation (see for example, Newey (1997) and Ahmad et al. (2005)), and this is the approach taken in this paper. The series method is known to be particularly suitable than kernel methods under certain type of restrictions, such as additive separability and varying coefficient structures, and it is also computational simpler because the results can be displayed by relatively few coefficients. Newey (1997) proposed a general series estimation approach for general nonparametric regression whilst Ahmad et al. (2005) proposed a series estimation procedure for a partially linear varying coefficient model in which $\theta(u) = u'\gamma$ in model (1).

The main contribution of this paper is of two folds. First, we consider series estimation procedure as an alternative and complement approach to the kernel smoothing technique to estimate the unknown functions in (1). Our estimation procedure is similar in spirit as Ahmad et al. (2005). In addition, we show that, via Monte Carlo simulation, the finite sample performance of the series estimator is as good as the two-step local linear kernel based estimator and it is simpler to compute. Second, we provide a synthesis on the asymptotic properties of the series estimator obtained in Newey (1997) and Ahmad et al. (2005).

The paper is organized as follows. Section 2 introduces the series estimation approach and develops the asymptotic properties of the proposed estimator. Section 3 reports some Monte Carlo simulation results. Section 4 concludes the paper. The proofs of the theorems are gathered in the appendix.

2. Series Estimation

To simplify the discussion and for exposition purposes, we confine our attention to the case in which u and z are one-dimensional. Extension to the multivariate u and z is straightforward and involves no fundamentally new ideas. However, implementation with both u and z having more than two dimensions may have difficulty due to the curse of dimensionality.

Let $\{y_i, u_i, x_i, z_i\}_{i=1}^n$ be an *i.i.d.* random sample from (1). Note that extension of our estimation method to dependent time series data is more complicated and beyond the scope of the paper, and hence we will not consider it here. With the series estimation method, we approximate unknown coefficient function $\theta(.)$ by some linear combination of L known base functions $\{p^L\}$ such that $\theta(u) = p^L(u)'\alpha$ where $p^L(u) = [p_1(u),\ldots,p_L(u)]'$ is a $(L\times 1)$ vector of base functions and $\alpha = (\alpha_1,\ldots,\alpha_L)'$ is a $(L\times 1)$ vector of unknown parameters. Similarly, for $j = 1,\ldots,d$, we approximate the unknown coefficient functions $\beta_j(.)$ by $q_j^{k_j}(z)'\gamma_j^{k_j}$, a linear combination of k_j base functions, where $q_j^{k_j}(z) = [q_{j1}(z),\ldots,q_{jk_j}(z)]'$ is a $(k_j \times 1)$ vector of the base function and $\gamma_j^{k_j} = (\gamma_{j1},\ldots,\gamma_{jk_j})'$ is a $(k_j \times 1)$ vector of unknown parameters. The approximating functions $p^L(u)$ and $q_j^{k_j}(z)$ have the property that, as L and k_j grow, there are linear combinations of $p^L(u)$ and $q_j^{k_j}(z)$ that can approximate respectively, any smooth function $\theta(u)$ and $\beta_j(z)$ arbitrarily well in mean square errors.

Define $K = \sum_{j=1}^d k_j$, following Ahmad et al. (2005), we use a linear combination of K functions, $q^K(x_i, z_i)'\gamma$ to approximate $x_i'\beta(z_i)$ where $q^K(x_i, z_i) = [x_{i1}q_1^{k_1}(z_i)',\ldots,x_{id}q_d^{k_d}(z_i)']'$ and $\gamma = (\gamma_1^{k_1'},\ldots,\gamma_d^{k_d'})'$ are $(K\times 1)$ matrices. For convenient, we now introduce some matrix notations. We define $(n\times 1)$ vector $Y = (y_1,\ldots,y_n)'$, an $(n\times 1)$ vector $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_n)'$, an $(n\times L)$ matrix $P = (p^L(u_1),\ldots,p^L(u_n))'$, an $(n\times K)$ matrix $Q = (q^K(x_1, z_1),\ldots,q^K(x_n, z_n))'$, $G = (x_1'\beta(z_1),\ldots,x_n'\beta(z_n))'$, and $\Theta = (\theta(u_1),\ldots,\theta(u_n))'$, then (1) can be rewritten as

$$Y = P\alpha + Q\gamma + \varepsilon, \tag{2}$$

where $\varepsilon = e + \{\Theta - P\alpha\} + \{G - Q\gamma\}$. Let $\hat{\alpha}$ and $\hat{\gamma}$ denote the least squares estimator of α and γ, respectively, obtaining by regressing Y on (P, Q). By using the standard argument of partitioned regressions, the estimators $\hat{\alpha}$ and $\hat{\gamma}$ are given by

$$\hat{\alpha}_L = (P'M_Q P)^- P'M_Q Y, \tag{3}$$

$$\hat{\gamma}_K = (Q'M_P Q)^- Q'M_P Y, \tag{4}$$

where $(.)^-$ denotes generalized inverse and we use the subscripts L and K to denote that these estimators are dependent of the number of approximating functions; $M_Q = I_n - Q(Q'Q)^- Q'$ and $M_P = I_n - P(P'P)^- P'$. However, under the conditions given below, both $(Q'Q)$ and $(P'P)$ will be asymptotically nonsingular matrices, hence the generalized inverses will be standard inverses for large n. Consequently, the estimators of $\theta(.)$ and $\beta_j(.)$ can be obtained, respectively as $\hat{\theta}(u) = P(u)\hat{\alpha}_L$ and $\hat{\beta}_j(z) = q_j^{k_j}(z)'\hat{\gamma}_j^{k_j}, \ j = 1,\ldots,d.$

Now we derive the convergence rates as well as the asymptotic normality of the proposed series estimators. To do this, we first define the following.

A function $f(u,x,z)$ is said to belong to the functional coefficient partially linear class of functions $\Im$ $(f \in \Im)$ if $f(u,x,z) = \delta(u) + \sum_{j=1}^{d} x_j \xi_j(z)$ for some continuous functions $\delta(u)$ and $\xi_j(z)$; $E[\delta(u)^2] < \infty$, $\sum_{j=1}^{d} E[x_j^2 \xi_j(z)^2] < \infty$ and $\delta(0) = 0$. Then, for any scalar or vector function $h(u,x,z)$, we use the notation $E_\Im[h(u,x,z)]$ to denote the projection of $h(u,x,z)$ onto additive functional coefficient functional space $\Im$ (under L_2-norm). Let $\wp(u,x,z) = h(u,x,z) - E_\Im[h(u,x,z)]$, then it follows that

$$E\{\wp(u,x,z)\wp(u,x,z)'\} = \inf_{f \in \Im} E\{[h(u,x,z) - f(u,x,z)][h(u,x,z) - f(u,x,z)]'\},$$

where infimum is in the sense that

$$E\{\wp(u,x,z)\wp(u,x,z)'\} \leq E\{[h(u,x,z) - f(u,x,z)][h(u,x,z) - f(u,x,z)]'\},$$

for all $f \in \Im$ and for square matrices A and B, $A \leq B$ indicates that $A - B$ is negative semidefinite.

4

Let Ω denote the $(L+K)\times(L+K)$ variance-covariance matrix of $(p^L(u_i)',q^K(x_i,z_i)')'$ whose smallest eigenvalue is bounded above zero, and the largest eigenvalue is bounded for every L and K, that is $\Omega = E\{[p^L(u_i),q^K(x_i,z_i)]'[(p^L(u_i),q^K(x_i,z_i)]'\}$. The matrix Ω can be decomposed as

$$\begin{pmatrix} \Omega_{pp} & \Omega_{pq} \\ \Omega_{qp} & \Omega_{qq} \end{pmatrix},$$

where $\Omega_{pq} = \Omega_{qp}'$. Thus, the conditional variance of $p^L(u_i)$ given $q^K(x_i,z_i)$ is given by $\Omega_{pp|q} = \Omega_{pp} - \Omega_{pq}\Omega_{qq}^{-1}\Omega_{qp}$, and the conditional variance of $q^K(x_i,z_i)$ given $p^L(u_i)$ is $\Omega_{qq|p} = \Omega_{qq} - \Omega_{pq}\Omega_{pp}^{-1}\Omega_{qp}$. We use "$\xrightarrow{d}$" to denote convergence in distribution, $||B|| = (B'B)^{1/2}$ if B is a vector and $||B|| = (tr(B'B))^{1/2}$ if B is a matrix where $tr(.)$ is the trace operator. The following assumptions will be used to establish the convergence rates as well as the asymptotic normality of $\hat{\theta}(u)$ and $\hat{\beta}_j(z)$, $j=1,\ldots,d$.

Assumption 1: (i) $\{y_i,u_i,x_i,z_i\}_{i=1}^n$ are independent and identically distributed as (y_1,u_1,x_1,z_1) and the support of (u_1,x_1,z_1) is a compact subset of $\mathfrak{R}^{d+2}$; $(ii)\,\mathrm{var}(y_1 \mid u_1,x_1,z_1) = \sigma^2(u,x,z)$ is a bounded function on the support of (u_1,x_1,z_1).

Assumption 2: For every L, there is a non-singular matrix A such that for $P^L(u) = Ap^L(u)$: (i) the smallest eigenvalue of $E[P^L(u_i)P^L(u_i)']$ is bounded away from zero uniformly in L; (ii) there exists a sequence of constants $\zeta_0(L)$ that satisfy the condition $\sup_{u\in S_u} || P^L(u) || \leq \zeta_0(L)$ where $L = L(n)$ is non-random such that $(\zeta_0(L)^2)L / n \to 0$ as $n \to \infty$, where S_u is the support of u_1.

Assumption 3: For every K, there is a nonsingular matrix B such that for $Q^K(x,z) = Bq^K(x,z)$: (i) the smallest eigenvalue of $E[Q^K(x_i,z_i)Q^K(x_i,z_i)']$ is bounded away from zero uniformly in K; (ii) there exists a sequence of constants $\xi_0(K)$ that satisfy the

5

condition $\sup_{(x,z)\in S_{(x,z)}} \| Q^K(x,z) \| \leq \xi_0(K)$ where $K = K(n)$ is non-random such that

$(\xi_0(K)^2)K / n \to 0$ as $n \to \infty$, where $S_{(x,z)}$ is the support of (x_1, z_1).

Assumption 4: There exists $\delta_0 > 0$ such that $\sup_{u\in S_u} |\theta(u) - p^L(u)'\alpha_L| = O(L^{-\delta_0})$ for every L.

Assumption 5: For $g(x,z) = \sum_{j=1}^{d} x_j \beta_j(z)$, there exist some $\delta_j > 0$ $(l = 1,...,d)$, $\gamma_g = \gamma_{gK} = (\gamma_1^{k_1'}, ..., \gamma_d^{k_d'})'$, such that $\sup_{(x,z)\in S_{(x,z)}} |g(x,z) - q^K(x,z)'\gamma_g| = O(\sum_{j=1}^{d} k_j^{-\delta_j})$. In addition, $\min\{k_1,...,k_d\} \to \infty, \sqrt{n}(\sum_{j=1}^{d} k_j^{-\delta_j}) \to 0$ as $n \to \infty$.

Most the above assumptions are adopted from Newey (1997) and Ahmad et al. (2005) for the purpose of our analysis. Assumption 1 is standard for series estimation with *i.i.d.* data, albeit the bounded conditional variance is difficult to relax without affecting the rates of convergence. Assumptions 2 and 3 impose bounded second moment matrices away from singularities and restricting the magnitudes of the series terms. Assumptions 4 and 5 state that there exists some positive constants such that the uniform approximation errors to the functions shrink at particular rates. Assumptions 4 and 5 are not the weakest conditions but it is known that many series functions satisfy these conditions (e.g. power series and splines).

Under the above assumptions, we can now state our main asymptotic results.

Theorem 1: Under Assumptions 1-5, as $n \to \infty$, we have

(i) $\int [\hat{\theta}(u) - \theta(u)]^2 dF_u(u) = O_p((L / n) + (K / n) + (L^{-2\delta_0} + \sum_{j=1}^{d} k_j^{-2\delta_j}))$ *where $F_u(u)$ is the cumulative distribution function of u.*

(ii) $\int [\hat{\beta}_j(z) - \beta_j(z)]^2 dF_z(z) = O_p((L / n) + (K / n) + (L^{-2\delta_0} + \sum_{j=1}^{d} k_j^{-2\delta_j})), \quad j = 1,...,d,$ *where $F_z(z)$ is the cumulative distribution function of z.*

Theorem 1 implies that the convergence rate of $\hat{\theta}(u)$ and $\hat{\beta}_j(z)$ $(j = 1,...,d)$ depends on both L and K, and it consists of two terms. The first term $((L / n) + (K / n))$ is essentially due to the

convergence rate of the variance whereas the second term $(L^{-2\delta_0} + \sum_{j=1}^{d} k_j^{-2\delta_j})$ corresponds to the convergence rate of the squared bias. The next theorem gives the asymptotic normality of $\hat{\theta}(z)$ and $\hat{\beta}(z) = (\hat{\beta}_1(z), ..., \hat{\beta}_d(z))'$.

Theorem 2: Under Assumptions 1-5, and in addition, $\sqrt{n}L^{-\delta_0} \to 0$, $\sqrt{n}(\sum_{j=1}^{d} k_j^{-2\delta_j}) \to 0$ then as $n \to \infty$, we have

(i) $V_u(u, L, K)^{-1/2}(\hat{\theta}(u) - \theta(u)) \xrightarrow{d} N(0,1)$,

(ii) $V_z(x, z, L, K)^{-1/2}(\hat{\beta}(z) - \beta(z)) \xrightarrow{d} N(0, I_n)$,

where $V_u(.) = n^{-1}\sigma^2(u, x, z)p^L(u)'\Omega_{pp|q}^{-1}p^L(u)$ *and* $V_z(.) = n^{-1}\sigma^2(u, x, z)q^K(x, z)'\Omega_{qq|p}^{-1}q^K(x, z)$. *Consistent estimators* $V_u(.)$ *and* $V_z(.)$ *are given by* $\hat{V}_u(.) = n^{-1}\hat{\sigma}^2(u, x, z)p^L(u)'\hat{\Omega}_{pp|q}^{-1}p^L(u)$ *and* $\hat{V}_z(.) = n^{-1}\hat{\sigma}^2(u, x, z)q^K(x, z)'\hat{\Omega}_{qq|p}^{-1}q^K(x, z)$ *where* $\hat{\Omega}_{pp|q}^{-1}$ *and* $\hat{\Omega}_{qq|p}^{-1}$ *are obtained from the partitioned matrices of* $\hat{\Omega} = n^{-1}\sum_{i=1}^{n}[p^L(u_i), q^K(x_i, z_i)][(p^L(u_i), q^K(x_i, z_i)]'$ *and* $\hat{\sigma}^2(u, x, z) = n^{-1}\sum_{i=1}^{n}\hat{e}'_i\hat{e}_i$ *where* $\hat{e}_i = y_i - p^L(u_i)'\hat{\alpha} - q^K(x_i, z_i)'\hat{\gamma}$.

The basic ideas of the proofs of both Theorems 1 and 2 are mainly from Newey (1997) and Ahmad et al. (2005). They are given in the Appendix for the readers' interest.

Note that the convergence rates of (i) and (ii) in Theorem 2 are not $\sqrt{n}$, and as in the case of nonparametric regression estimator, they are slower than $\sqrt{n}$ as the smoothing parameters shrink. The convergence rates are implicitly embedded in the variance terms $V_u(.)$ and $V_z(.)$. In addition, Theorem 2 shows how to construct the asymptotic standard errors as well as the confidence intervals for $\theta(z)$ and $\beta(z)$.

Remarks:

(i) The results presented in Theorems 1 and 2 provide a synthesis on the convergence rates and asymptotic normality of series estimator derived in Newey (1997) and Ahmad et al. (2005). For example, if we set $\theta(u) = u'\gamma = 0$ for all u in (1) the results in Theorem 1 reduces to the results of Ahmad et al. (2005). Similarly, if we set the general function considered in Newey (1997) to a varying coefficient model such as (1), the results in Theorem 2 can also be deduced from the results presented in Newey (1997).

(ii) Theorem 1 gives the convergence rates of the series estimator for $\theta(u)$ and $\beta_j(z)$ and they simultaneously depend on the approximating terms L and K, hence it is more difficult to determine the optimal choice for L and K that balance between biases and variances. However, in a special case where we set $k_j = L$ for all j then it is easy to show that the optimal choice for L such that $(L / n) + (dL / n) = (d+1)L^{-2\delta_0}$ is $L^* = n^{1/(1+2\delta_0)}$ for $\delta_0 > 0$. Moreover, in order to satisfy the asymptotic normality in Theorem 2, we would need to choose larger L than the optimal L^* (i.e., undersmoothing) since the condition $\sqrt{n}L^{-\delta_0} \to 0$ also needs to hold. However, if one wishes not to undersmoothing data, the asymptotic normality of the estimator, in general, can still be achieved but the estimator has the asymptotic bias component which is unknown in general, and equals to the approximation errors, see for example, Hansen (2013, Chapters 12.12 - 12.14), and Huang (2003) for more detailed discussion on the difficulty of obtaining asymptotic bias for general case with splines.

3. Monte Carlo Simulation

To examine the finite sample performance of the proposed series estimator, and also to compare it with the two-step local linear kernel based estimator suggested by Wong et al. (2008), we conduct some simulations. To this end, we consider the same the data generating process (DGP) as in Wong et al. (2008). The model is:

$$Y_i = \theta(u_i) + x_{1i}\beta_1(z_i) + x_{2i}\beta_2(z_i) + \varepsilon_i,$$

where $\theta(u) = u + \exp(-16u^2)$, $\beta_1(z) = 0.138 + (0.316 + 0.982z)\exp(-3.89z^2)$ and

$\beta_2(z) = -0.437 - (0.659 + 1.260z)\exp(-3.89z^2)$. The error term ε_i are generated as

$\varepsilon_i \sim i.i.d.\, N(0,0.2^2)$, u_i, z_i, x_{1i} and x_{2i} are each generated by the $i.i.d.\, Uniform[-5,5]$ distribution. The sample size is $n = 400$ and the number of replications is 400. We compute and compare the estimated square root of mean square errors (RMSE) of $\hat{\theta}(u_i)$ and $\hat{\beta}_j(z_i)$ which respectively defined by $RMSE(\hat{\theta}(.)) = (1/R)\sum_{r=1}^{R}\left[(1/n)\sum_{i=1}^{n}(\hat{\theta}_r(u_i) - \theta(u_i))^2\right]^{1/2}$ and

$RMSE(\hat{\beta}_j(.)) = (1/R)\sum_{r=1}^{R}\left[(1/n)\sum_{i=1}^{n}(\hat{\beta}_{jr}(z_i) - \beta_j(z_i))^2\right]^{1/2}$ for $j = 1,2$, where R is the number of replications, $\hat{\theta}(u)$ and $\hat{\beta}_j(z)$ are respectively, the estimates of $\theta(u)$ and $\beta_j(z)$ from the rth replication based on either the series method or the two-step local linear approach. For the series method, we use a univariate cubic B-spline which defined as

$$B(z \mid t_0,\ldots,t_4) = \frac{1}{3!}\sum_{j=0}^{4}(-1)^j\binom{4}{j}[\max(0, z - t_j)]^3,$$

where $(t_0,\ldots,t_4)$ are equally space design knots. For the two-step local linear and one-step back fitting approach with optimal choice of the smoothing parameters h_1 and h_2, see Wong et al. (2008). For the series method, we select the number of approximating terms L and K by the leave-one-out least squares cross-validation (CV) that minimize the approximate MSE. Recently, Hansen (2012) shows that for estimating regression function with series method, the CV approach is not only computationally simple method, it is also asymptotically optimal among the alternative selection criteria, in the sense that the CV-selected estimator is asymptotically equivalent to the infeasible best-fitting estimator when all approaches were evaluated based on integrated MSE.

The simulation results are displayed in Table 1. Note that, the last column in Table 1 shows the sum of the three RMSE. From Table 1, we observe that the series and the two-step kernel methods give similar estimation results for the three functional coefficients. Consequently, both the series and two-step kernel methods can be a useful tool in estimating the FCPLR model.

Next, we evaluate the 95% empirical coverage probabilities of the functional coefficients based on Theorem 2. Specifically, for a given value of u and z, the 95% confident intervals for $\theta(u)$

and $\beta_j(z)$, $j = 1, 2$, are computed as $\hat{\theta}(u) \pm 2\hat{s}_u$ where $\hat{s}_u = \sqrt{\hat{V}_u}$ and $\hat{\beta}_j \pm 2\hat{s}_z^{jj}$ where

$\hat{s}_z^{jj} = \sqrt{\hat{V}_z^{jj}}$ with $\hat{V}_z^{jj}$ is the jth diagonal element of the matrix $\hat{V}_z$. For simplicity, we evaluate these quantities at three different quantiles of u and z: 25%, 50% and 75%. In addition, since undersmoothing is required, we first assume $L = K$ for all cases, and use $L = (1.5L^*, 2.5L^*)$ as the number of approximating terms, where L^* is the optimal value selected by using CV procedure. The results are displayed in Table 2. Our results indicated that the coverage probabilities of interval estimates based on the series estimators are very close to the true coverage probabilities and there are very little distortions.

4. Concluding Remark

In this paper, we propose an alternative and complementing approach for estimating the FCPLR model. Specially, we suggest a series method which is simpler to implement in practice than the kernel method. We established a synthesis on the consistency and asymptotic normality of the proposed estimator. Limited Monte Carlo simulations suggested the proposed estimator perform well in finite sample.

Note that, in practical application where we have dependent time series and this may not fit the general setting of this paper where the data are assumed to be independent. For dependent time series data, we conjecture that our approach can be still be applied to geometrically β-mixing and strictly stationary data; and under a different set of regularity conditions (especially on the boundedness conditions), the convergence rate of the estimator still hold, although the asymptotic variance may have a different forms. Extension our approach to dependent time series data and perhaps dynamic panel data (see Lee (2013) and Tran (2014)) are interesting problems and deserves further research.

References

Ahmad, I., S. Leelahanon and Q. Li (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *Ann. Statist.* 33: 258-283.

Fan, J. and W. Zhang (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface.* 1: 179-195.

Hansen, B.E. (2012). Nonparametric sieve regression: least squares, averaging least squares, and cross-validation. *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Chapter 8: 215-248.

Hansen, B.E. (2013). Econometrics. *Mimeo*, Department of Economics, University of Wisconsin, Madison WI.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* 31: 1600-1635.

Lee, Y. (2013). Nonparametric estimation of dynamic panel models with fixed effect. *Econometric Theory.* 30: 1315-1347.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *J. of Econometrics.* 79: 147-168.

Tran, K.C. (2014). Nonparametric estimation of functional-coefficient partially linear dynamic panel data models with fixed effects. *Economics Bulletin*. 34(3): 1751-1761.

Wong, H., R. Zhang, W-C Ip and G. Li (2008). Functional-coefficient partially linear regression model. *J. of Mult. Anal.* 99: 278-305.

Appendix A: Mathematical Proofs

First we present some useful lemmas which are the results that will be used in the proofs of Theorem1 and Theorem 2. Following the arguments of Newey (1997), we will assume without loss of generality, that $A = I_L$ and $B = I_K$ where I_R is an identity matrix of dimension R; A and B are defined in Assumptions 2 and 3. Thus, $P^L(u) = p^L(u)$, $Q^K(x,z) = q^K(x,z)$ and $\Gamma_L = E[p^L(u)p^L(u)'] = I_L$, $\Gamma_K = E[q^K(x,z)q^K(x,z)'] = I_K$. Also, we define an indicator function 1_n which equals to 1 if $(W'W)$ is nonsingular and 0 otherwise, for $W = P$ or Q.

Lemma A.1 (i) $\left\| \hat{\Gamma}_L - I_L \right\| = O_p(\zeta_0(L)\sqrt{L/n}) = o_p(1)$;

(ii) $\left\| \hat{\Gamma}_K - I_K \right\| = O_p(\zeta_0(K)\sqrt{K/n}) = o_p(1)$

where $\hat{\Gamma}_L = P'P/n$ and $\hat{\Gamma}_K = Q'Q/n$.

PROOF: See Theorem 1 of Newey (1997).   □

Lemma A.2 (i) $\left\| \tilde{\alpha} - \alpha \right\|^2 = O_p(L^{-2\delta_0})$ *where* $\tilde{\alpha} = (P'P)^{-1}P'\Theta$, α *satisfies Assumption 4.*

(ii) $\left\| \tilde{\gamma}_g - \gamma_g \right\|^2 = O_p(\sum_{j=1}^{d} k_j^{-2\delta_j})$ *where* $\tilde{\gamma} = (Q'Q)^{-1}Q'G$, γ_g *satisfies Assumption 5.*

PROOF: The proof follows the same arguments as in the proof of Lemma A.2 in Ahmad et al. (2005).   □

Lemma A.3 (i) $1_n \left\| (P'P/n)^{-1}(P'e/n) \right\|^2 = O_p(L/n)$

(ii) $1_n \left\| (Q'Q/n)^{-1}(Q'e/n) \right\|^2 = O_p(K/n)$

PROOF: To prove (i), taking the conditional expectation of (i), we have

$$E[1_n \left\| (P'P/n)^{-1}(P'e/n) \right\|^2 \mid u, x, z]$$

$$= 1_n E\{[(e'P/n)(P'P/n)^{-1}(P'P/n)^{-1}(P'e/n)] \mid u, x, z\}$$

$$\leq O_p(1)1_n \, tr\,(P(P'P)^{-1}P'E(ee' \mid u, x, z)/n)$$

$$\leq O_p(1)1_n C(L/n)$$

by Lemma A.1 and Assumption 1. Thus, $1_n \left\| (P'P/n)^{-1}(P'e/n) \right\|^2 = O_p(L/n)$. The proof of (ii) follows similarly. $\square$

Lemma A.4 $\left\| \hat{\Omega} - \Omega \right\| \to_p 0$ where $\hat{\Omega} = n^{-1}\sum_{i=1}^{n}[p^L(u_i), q^K(x_i, z_i)][(p^L(u_i), q^K(x_i, z_i)]'$

PROOF: First note that by Assumption 6 and by using argument as in Newey (1997), without loss of generality we set $\Omega = I$ where I is an identity matrix of dimension $(L+K)$. Then it can be shown that as in Lemma 1, $\left\| \hat{\Omega} - \Omega \right\| = O_p \, \zeta_0(L+K)\sqrt{(L+K)/n} = o_p(1)$. $\square$

Proof of Theorem 1: The basic idea of the proof of this theorem is mainly from Newey (1997) and Ahmad et al. (2005).

(i) By (2) and (3), we can write

$$\hat{\alpha} = (P'P)^{-1}P'(Y - Q\hat{\gamma})$$

$$= (P'P)^{-1}P'(P\alpha + Q\gamma + e + (\Theta - P\alpha) + (G - Q\gamma) - Q\hat{\gamma})$$

$$\hat{\alpha} - \alpha = (P'P/n)^{-1}(P'e/n) + (P'P/n)^{-1}(P'(\Theta - P\alpha)/n) + (P'P/n)^{-1}(P'(G - Q\gamma)/n) -$$

$$(P'P/n)^{-1}(P'Q(\hat{\gamma} - \gamma)/n)$$

Hence,

13

$$1_n \left\| \hat{\alpha} - \alpha \right\|^2 \le 1_n \left\| (P'P / n)^{-1}(P'e / n) \right\|^2 + \left\| (P'P / n)^{-1}(P'(\Theta - P\alpha) / n) \right\|^2 +$$

$$1_n \left\| (P'P / n)^{-1}(P'(G - Q\gamma) / n) \right\|^2 +$$

$$1_n \left\| (P'P / n)^{-1}(P'Q(\hat{\gamma} - \gamma) / n) \right\|^2$$

The first term $1_n \left\| (P'P / n)^{-1}(P'e / n) \right\|^2 = O_p(L / n)$ by Lemma A.3(i). The second term

$1_n \left\| (P'P / n)^{-1}(P'(\Theta - P\alpha) / n) \right\|^2 = O_p(L^{-2\delta_0})$ by Lemma A.2(i). Similarly, it is

straightforward to show that the third term $1_n \left\| (P'P / n)^{-1}(P'(G - Q\gamma) / n) \right\|^2 = O_p(\sum_{j=1}^d k_j^{-2\delta_j})$

because the function G is approximated by $Q\gamma$ and hence, the convergence rate will depend on

$K = \sum_{j=1}^d k_j$. The last term looks more complicated because it involves the convergence rate of

$(\hat{\gamma} - \gamma)$. However from (4), $(\hat{\gamma} - \gamma)$ can be expressed explicitly as a function of $(\hat{\alpha} - \alpha)$ and by

substituting this expression into the last term and solve for $(\hat{\alpha} - \alpha)$, it can be shown that

$1_n \left\| (P'P / n)^{-1}(P'Q(\hat{\gamma} - \gamma) / n) \right\|^2 = O_p(K / n)$. Thus, by combining the above results, also by

noting that $1_n \to 1$ almost surely, we have

$$\left\| \hat{\alpha} - \alpha \right\|^2 = O_p((L / n) + (K / n) + (L^{-2\delta_0} + \sum_{j=1}^d k_j^{-2\delta_j})) \tag{A.1}$$

Next, by triangle inequality, we have

$$\int [\hat{\theta}(u) - \theta(u)]^2 dF_u(u) = \int [p^L(u)'(\hat{\alpha} - \alpha) + (p^L(u)'\alpha - \theta(u))]^2 dF_u(u)$$

$$\le \left\| \hat{\alpha} - \alpha \right\|^2 + \int [p^L(u)'\alpha - \theta(u)]^2 dF_u(u)$$

$$= O_p((L / n) + (K / n) + (L^{-2\delta_0} + \sum_{j=1}^d k_j^{-2\delta_j})) + O(L^{-2\delta_0})$$

$$= O_p((L / n) + (K / n) + (L^{-2\delta_0} + \sum_{j=1}^d k_j^{-2\delta_j}))$$

by (A.1), Lemma A.1 and Assumption $4(i)$. Thus, we have proved Theorem $1(i)$. The proof of Theorem $1(ii)$ follows the same arguments as in the proof above and hence omitted here. $\square$

Proof of Theorem 2: The detailed proof of this theorem is a straightforward extension of the Theorems 2 and 3 of Newey (1997), so we discuss the heuristic idea of the proof here. First observe that

$$
\begin{aligned}
1_n \sqrt{n}(\hat{\theta}(u) - \theta(u)) &= 1_n \sqrt{n} P(\hat{\alpha} - \alpha) \\
&= 1_n \sqrt{n} P\{(P'M_Q P)^{-1} P'M_Q(\Theta - P\alpha)\} + \\
& \qquad 1_n \sqrt{n} P\{(P'M_Q P)^{-1} P'M_Q(G - Q\gamma)\} + \\
& \qquad 1_n \sqrt{n} P\{(P'M_Q P)^{-1} P'M_Q e\}
\end{aligned}
\tag{A.2}
$$

and

$$
\begin{aligned}
1_n \sqrt{n}(\hat{\beta}(z) - \beta(z)) &= 1_n \sqrt{n} Q(\hat{\gamma} - \gamma) \\
&= 1_n \sqrt{n} Q\{(Q'M_P Q)^{-1} Q'M_P(G - Q\gamma)\} + \\
& \qquad 1_n \sqrt{n} Q\{(Q'M_P Q)^{-1} Q'M_P(\Theta - P\alpha)\} + \\
& \qquad 1_n \sqrt{n} Q\{(Q'M_P Q)^{-1} Q'M_P e\}
\end{aligned}
\tag{A.3}
$$

By Lemma A.4, we have $\left\|\hat{\Omega} - \Omega\right\| \to_p 0$ as $n \to \infty$, thus, the first two terms in $(A.2)$ and $(A.3)$ can be shown to be asymptotically negligible by Assumptions A.4 and A.5. By combining the last term in $(A.2)$ and $(A.3)$ along with the result of partitioned regression, yields

$$
\begin{pmatrix} 1_n \sqrt{n} P\{(P'M_Q P)^{-1} P'M_Q e\} \\ 1_n \sqrt{n} Q\{(Q'M_P Q)^{-1} Q'M_P e\} \end{pmatrix} = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \hat{\Omega}^{-1} \begin{pmatrix} 1_n P'e / \sqrt{n} \\ 1_n Q'e / \sqrt{n} \end{pmatrix}
$$

First note that $\left\|\hat{\Omega}^{-1} - \Omega^{-1}\right\| \to_p 0$ as $n \to \infty$. Second, following Newey (1997, proof of Theorem 3) and the fact that $1_n \to 1$ almost surely, the limit distribution of the quantity $\begin{pmatrix} P'e / \sqrt{n} \\ Q'e / \sqrt{n} \end{pmatrix}$ is approximately normal with mean zero and variance $\sigma^2(x, u, z)\Omega$. By using the inverse matrix formula of the partitioned matrix, we have

$$
\Omega^{-1} = \begin{pmatrix} \Omega_{pp} & \Omega_{pq} \\ \Omega_{qp} & \Omega_{qq} \end{pmatrix}^{-1} = \begin{pmatrix} \Omega_{pp|q}^{-1} & -\Omega_{pp|q}^{-1}\Omega_{pq}\Omega_{qq}^{-1} \\ -\Omega_{qq}^{-1}\Omega_{qp}\Omega_{pp|q}^{-1} & \Omega_{qq}^{-1} + \Omega_{qq}^{-1}\Omega_{qp}\Omega_{pp|q}^{-1}\Omega_{pq}\Omega_{qq}^{-1} \end{pmatrix}
$$

$$
= \begin{pmatrix} \Omega_{pp}^{-1} + \Omega_{pp}^{-1}\Omega_{pq}\Omega_{qq|p}^{-1}\Omega_{qp}\Omega_{pp}^{-1} & -\Omega_{pp}^{-1}\Omega_{pq}\Omega_{qq|p}^{-1} \\ -\Omega_{qq|p}^{-1}\Omega_{qp}\Omega_{pp}^{-1} & \Omega_{qq|p}^{-1} \end{pmatrix}
$$

and the desired result is obtained using the above expression. $\square$

Table 1: RMSE of Functional Coefficients

Methods	$\theta(u)$	$\beta_1(z)$	$\beta_2(z)$	Total
2-Step Kernel	0.00251	0.00732	0.00701	0.01684
Series	0.00250	0.00733	0.00700	0.01683

Table 2: 95% Coverage Probabilities

Quantile of u or z	$\theta(u)$		$\beta_1(z)$		$\beta_2(z)$	
	$L=1.5L^*$	$L=2.5L^*$	$L=1.5L^*$	$L=2.5L^*$	$L=1.5L^*$	$L=2.5L^*$
0.25	0.952	0.951	0.951	0.950	0.953	0.952
0.50	0.957	0.950	0.954	0.955	0.953	0.952
0.75	0.952	0.953	0.951	0.953	0.953	0.956