Piecewise Local Linear Estimation of Functional Equilibrium Relationships

Anurag Banerjee
Durham University
Department of Economics and Finance
a.n.banerjee@durham.ac.uk

Jean-Yves Pitarakis
University of Southampton
Economics Division
j.pitarakis@soton.ac.uk

2011

Abstract

We introduce a new nonparametric approach for estimating a simple varying coefficient model with a unit root nonstationarity. Our method is based on a piecewise local least squares principle and is computationally simple to implement. We establish its asymptotic properties and evaluate its performance in finite samples. Our working model also allows us to formalise the concept of a long run functional equilibrium relationship analogous to the well known cointegration concept within a constant coefficient setting.

Keywords: Functional Coefficients, Unit Roots, Cointegration, Piecewise Local Linear Linear Estimation.

JEL: C22, C50.

1 Introduction

A vast body of research in the recent time series econometrics literature has concentrated on developing methods of capturing nonlinear regime specific behaviour in the joint dynamics linking economic and financial variables. An important complication that arises when moving from simple linear structures with constant coefficients to such models with nonlinear dynamics has to do with the open ended nature of the functional forms one may want to adopt for describing the changing nature of the model parameters and underlying moments. Popular parametric specifications include the well known threshold models, Markov switching models, models with structural breaks among numerous others. Although such models can allow researchers to capture rich and economically meaningful nonlinearities the ad-hoc nature of the functional forms may also be seen as problematic. An alternative to having to take a stand on a particular functional form is to instead allow the changing coefficients to be described by some unknown function to be estimated from the data as for instance in $y = f(q)x + e$. Such semiparametric specifications are commonly referred to as varying or functional coefficient models and were introduced in the early work of Cleveland, Grosse and Shyu (1991), Hastie and Tibshirani (1993), Chen and Tsay (1993), Fan and Zhang (1999) amongst numerous others (see also Fan and Yao (2003) and references therein). An important motivation underlying this class of models is their ability to capture rich dynamics in a flexible way while at the same time avoiding the curse of dimensionality characterising fully nonparametric specifications.

The most common way of estimating the unknown functions of such varying coefficient models is through kernel smoothing and local polynomial techniques. These typically reduce to a weighted least squares type of objective function with the weights dictated by some chosen kernel function (see Fan and Gijbels (1992, 1996)). Our objective in this paper is to propose an alternative estimation approach based on a piecewise linear least squares principle and to obtain its properties within a nonstandard context that allows for the presence of a unit root variable as in the recent work of Juhl (2009), Xiao (2009) and Cai, Li and Park (2009). Our method is different from kernel smoothing based methods, does not generally require the differentiability of the density of q and is shown to have good finite sample properties.

Our varying coefficient specification also allows us to define a novel concept which we refer to as *functional cointegration* analogous to the well known linear cointegration property that arises in specifications linking I(1) variables linearly. In this sense our work also falls within the bounds of the very recent literature on nonlinear cointegration tackled from a purely nonparametric point of view (Karlsten, Myklebust and Tjostheim (2007), Wang and Phillips (2009), Kasparis and Phillips (2009) amongst others). Note that the idea of a nonlinear long run equilibrium relationship (attractor) was also put forward in the early work of Granger and Hallman (1989), Breitung (2001), Saikkonen and Choi (2004) amongst others.

The plan of the paper is as follows. Section 2 introduces and motivates our model. Section

3 describes our estimation methodology and derives its asymptotic properties. Section 4 explores its performance and finite sample properties. Section 5 concludes. All proofs are relegated to the appendix.

2 The Model and Motivations

We consider the following functional coefficient regression model

$$y_t = f_0(q_{t-d}) + f_1(q_{t-d})x_t + u_t \quad (1)$$

$$x_t = x_{t-1} + v_t \quad (2)$$

where u_t and v_t are stationary disturbance terms and $f_0(q_{t-d})$ and $f_1(q_{t-d})$ are unknown functions of the stationary scalar random variable q_{t-d} while x_t is taken as an I(1) process throughout. The particular choice of d is not essential for our analysis and will be set at $d = 1$ throughout. The generality of (1)-(2) can be seen by noting that it can easily be specialised to well known parametric specifications such as threshold effects as in $f_i(q_{t-1}) = \beta_{i1}I(q_{t-1} \leq \gamma)+\beta_{i2}I(q_{t-1} > \gamma)$ (see Gonzalo and Pitarakis (2006)) or ESTAR/LSTAR type of variants such as $f_i(q_{t-1}) = [1 + e^{-\gamma_i(q_{t-1}-c_i)}]^{-1}$ amongst others.

Before proceeding with our main goal of estimating the unknown functions $f_0(q)$ and $f_1(q)$ it is important to motivate our model in (1)-(2) as a long run equilibrium relationship. As it stands (1) cannot be interpreted as a stationary nonlinear combination of I(1) variables in a traditional sense. Indeed, it is easy to see that although x_t is a standard $I(1)$ process, y_t can no longer be viewed as I(1) as it would have been the case for instance if $f_0(q)$ and $f_1(q)$ were constants. Differently put, the concept of integratedness of order 0 or 1 is mainly relevant within a linear framework while not being very helpful when dealing with nonlinear transformations of variables. In the context of our model in (1) for instance it is straightforward to see that first differencing y_t will not result in a stationary process because of the time varying nature of the functional coefficients.

To gain further insight into this phenomenon consider a simplified version of (1) which we compactly write as $y_t = f_t x_t + u_t$ and with f_t denoting some stationary process. It is now clear that $\Delta y_t = f_t \Delta x_t + x_{t-1}\Delta f_t + \Delta u_t$ making it difficult to view Δy_t as a stationary process due to the presence of the term $x_{t-1}\Delta f_t$ which has a variance that grows with t. Instead, cointegration in the context of (1) is understood in the sense that although y_t and x_t have variances that grow with t, the *functional combination* given by u_t is stationary.

Because of these conceptual difficulties and for the purpose of motivating (1)-(2) we propose to use the concept of *Summability* as an alternative to the concept of I(1)'ness and which was proposed in Gonzalo and Pitarakis (2006) and more recently refined and formalised in Berenguer (2010) and Berenguer and Gonzalo (2011). A time series y_t is said to be summable of order δ, symbolically represented as $S_y(\delta)$, if the sum $S_y = \sum_{t=1}^{T}(y_t - d_t)$ is such that $S_y/T^{\frac{1}{2}+\delta} = O_p(1)$ as $T \to \infty$ and where d_t denotes a deterministic sequence. Note that in the context of this definition, a process that is I(d) can be referred to as $S_y(d)$ and the functional process introduced in (1) is clearly $S_y(1)$ as discussed further below. Using this concept of summability of order δ we can now provide a formal definition of the concept of functional cointegration as follows

Definition (Functional Cointegration) *Let y_t and x_t be $S_y(\delta_1)$ and $S_y(\delta_2)$ respectively. They are functionally cointegrated if there exists a functional combination $(1, -f_1(q_{t-1}))$ such that $z_t = y_t - f_1(q_{t-1})x_t$ is $S_y(\delta_0)$ with $\delta_0 < \min(\delta_1, \delta_2)$.*

Given the formal definition of functional cointegration presented above it is now clear that within our specification in (1), y_t and x_t are functionally cointegrated with $\delta_0 = 0$ and $\delta_1 = \delta_2 = 1$. This follows from the fact that taking u_t and q_t to be stationary processes ensures that $\sum y_t/T^{3/2} = O_p(1)$ while u_t is such that $\sum u_t/\sqrt{T} = O_p(1)$ as clarified further below. It is also worth highlighting the fact that within our specification in (1) we have $z_t = f_0(q_{t-1}) + u_t$ which is of the same order of magnitude as u_t since under our assumptions we will have $\sum f_0(q_{t-1})/T \xrightarrow{p} E[f_0(q_{t-1})]$ and $\sum f_0(q_{t-1})/T^{3/2} = o_p(1)$.

Having provided a rationale for our specification in (1)-(2) we next concentrate on the main goal of this paper which involves obtaining reliable estimates of the unknown functional coefficients $f_0(q)$ and $f_1(q)$ and exploring their asymptotic and finite sample properties. For this purpose we introduce a piecewise linear estimation approach as developed in Banerjee (1994, 2007) in the context of average derivative estimation and adapt it to the nonstationary functional coefficient setting given by (1)-(2). This will also allow us to compare our approach with the more commonly used kernel smoothing approaches.

3 Piecewise Local Linear Estimation

We now concentrate on the estimation of the unknown functional coefficients linking y_t and x_t. We propose to do that through a piecewise local linear procedure recently used in Banerjee (1994, 2007) in the context of average derivative estimation. With q_{t-1} denoting the argument of the functions of interest, we divide its support into k disjoint bins of equal length $|H_r| = h$, $r = 1, \ldots, k$ (note that q_{t-1} is not sorted in any particular order). For every q_{t-1} falling in the r^{th} bin we then fit the least squares line $y_t = \beta_{0r} + \beta_{1r}x_t + u_t$ connecting the $\{y_t, x_t\}$ data within the bin. More specifically, letting $\tilde{x}_t = (1, x_t)$ and $I_r(q_{t-1}) \equiv I(q_{t-1} \in H_r) = 1$ if q_{t-1} falls within the r^{th} bin and zero otherwise and $\beta_r = (\beta_{0r}, \beta_{1r})'$ we write

$$\hat{\beta}_r = S_{xx}^{(r)-1} S_{xy}^{(r)} \tag{3}$$

where $S_{xx}^{(r)} = \sum_{t=1}^{T} \tilde{x}_t \tilde{x}_t' I_r(q_{t-1})$ and $S_{xy}^{(r)} = \sum_{t=1}^{T} \tilde{x}_t y_t I_r(q_{t-1})$. Note that $\hat{\beta}_r$ provides the least squares estimators of the intercept and slope parameters characterising the linear regression line within each bin. Interestingly, in a series of recent papers, Senturk and Mueller (2005, 2006) also used an estimation technique similar to what we consider below in an unobserved variable setting under iid'ness and in which observed and unobserved variables are linked through functional coefficients.

Once the $\hat{\beta}_r$'s have been estimated within each bin, our estimator of the functional coefficients is then given by

$$(\hat{f}_0(q), \hat{f}_1(q)) = \left(\sum_{r=1}^{k} \hat{\beta}_{0r} I_r(q), \sum_{r=1}^{k} \hat{\beta}_{1r} I_r(q)\right) \tag{4}$$

with $I_r(q) = I(q \in H_r)$.

Having introduced the mechanics behind our estimator our main goal is to establish its asymptotic properties. We will initially be interested in the consistency properties of $\hat{\beta}_r$ and subsequently focus on obtaining a convenient expression for its limiting distribution. Since in this nonstationary setting consistency typically holds under minimally restrictive assumptions that can accomodate serial correlation and/or endogeneity we proceed and operate under a broad set of high level assumptions which we will subsequently specialise as we explore distributional properties. The following baseline assumptions will be maintained throughout the entire paper.

Assumptions A. *(i)* $q_t = \mu_q + u_{qt}$ *is such that* u_{qt} *is a strictly stationary, ergodic and strong mixing sequence with mixing numbers* α_m *satisfying* $\sum_{m=1}^{\infty} \alpha^{\frac{1}{m}-\frac{1}{r}} < \infty$ *for some* $r > 2$. *(ii) The density of* q *denoted* $g_q(q)$ *is strictly positive and satisfies* $\sup_q g_q(q) < \infty$, *(iii) the functional coefficients* $f_i(.)$ $i = 0, 1$ *are twice continuously differentiable in* q.

Assumptions A(i)-(iii) above impose a standard set of restrictions on our functional coefficients and their argument q_{t-1}. The differentiability of the $f_i(q_{t-1})'s$ will allow us to use their local Taylor expansions at a point q within each bin. Assumption A(i) also requires q_t to be stationary while allowing it to accomodate a very rich set of dynamics such as ARMA specifications. Since our estimation methodology requires fitting a least squares line within each bin of length $|H_r| = h$ it is also understood throughout this paper that for estimability purposes there are enough observations within each bin.

We next introduce a set of high-level assumptions that restrict the large sample behaviour of the innovations driving (1)-(2).

Assumption B. *As* $T \to \infty$ *and* $\sqrt{Th} \to \infty$ *with* $h \to 0$ *a multivariate invariance principle holds for* $(u_t, u_t I_{rt-1}, v_t)$. *We write* $(\sum_{t=1}^{[Ts]} u_t/\sqrt{T}, \sum_{t=1}^{[Ts]} u_t I_{rt-1}/\sqrt{Th}, \sum_{t=1}^{[Ts]} v_t/\sqrt{T}) \Rightarrow BM(\Omega) \equiv (B_u(s), B_{ur}(s), B_v(s))$ *a three dimensional Brownian Motion with long run covariance* $\Omega > 0$.

Note that our Assumption B does not explictly impose any restrictions on the interactions between u_t, v_t and I_{rt-1} beyond typical existence of moments and memory restriction requirements that ensure an invariance principle holds. Both u_t and v_t may be individually serially correlated and also have nonzero covariances at all leads and lags. Assumption B would hold for instance if $w_t = (u_t, u_t I_{rt-1}, v_t)'$ was taken to be strictly stationary, ergodic and strong mixing with mixing coefficients as in A(i) and additional existence of moment requirements such as $E|w_{it}|^{2+\rho} < \infty$ for some $\rho > 0$.

We are now in a position to state our main result which establishes the consistency of our piecewise local linear estimator. It is summarised in the following Proposition.

Proposition 1. *Under Assumptions A and B, as $T \to \infty$ and if $Th \to \infty$ and $Th^{3/2} \to 0$ as $h \to 0$ we have $(\hat{f}_0(q) - f_0(q)) = O_p(1/\sqrt{Th})$ and $(\hat{f}_1(q) - f_1(q)) = O_p(1/T\sqrt{h})$.*

The above proposition focused on the consistency of our proposed estimator under a setting that allows a great degree of generality in the dynamics linking (1) and (2). We note that the slope function converges at a faster rate than the intercept function (i.e. $T\sqrt{h}$ versus $\sqrt{Th}$). This is directly analogous to the standard linear cointegration setting in which the slope converges at rate T while the intercept converges at the slower $\sqrt{T}$ rate. Our convergence rates conform with related studies that explored the use of functional coefficients in unit root settings using kernel smoothing techniques (Juhl (2006), Xiao (2009), Cai et al. (2009)).

At this stage it is also useful to highlight the bias term that characterises the asymptotics of $\hat{\beta}_{1r}$ and that vanishes under $Th^{3/2} \to 0$. From expressions (17)-(18) in our proof of Proposition 1 it is straightforward to observe that

$$
\begin{aligned}
\hat{\beta}_{1r} &= f_1(q) + f_1'(q)\frac{E[(q_{t-1} - q)I_{rt-1}]}{E[I_{rt-1}]} + O_p\left(\frac{1}{T\sqrt{h}}\right) \\
&= f_1(q) + o(h) + O_p\left(\frac{1}{T\sqrt{h}}\right). \qquad (5)
\end{aligned}
$$

Note that the bias term $f_1'(q)E[(q_{t-1} - q)I_{rt-1}]/E[I_{rt-1}]$ is the result of using a first order Taylor expansion of $f_1(q_{t-1})$ around a $q \in H_r$ (see (15)). It is also interesting to note that the bias term that would arise if we were to use higher order expansions (say up to order m) is given by

$$
B_m = \frac{1}{E[I_{rt-1}]}\sum_{j=1}^{m}\frac{1}{j!}f^{(m-j)}(q)[E(q_{t-1} - q)^j I_{rt-1}] \qquad (6)
$$

and does not depend on derivatives of the pdf of q_{t-1} (e.g. $g_q^{(m-j)}$) as it is typically the case in kernel smoothing based methods.

Having established the consistency of our proposed estimators of $f_0(.)$ and $f_1(.)$ we next concentrate on explicit representations of their limiting distributions. To achieve this, Assumption C below imposes further assumptions more explicitly restricting the dynamics of the sequence $u_t I_{rt-1}$ and its interaction with the v_t series. Letting $\eta_{rt} = u_t I_{rt-1}/\sqrt{h}$ and $\gamma_\ell(q) = E[u_{t+\ell}v_t|q_{t+\ell-1}]$ we require the following to hold.

Assumption C. *(i) The sequence $\{\eta_{rt}\}$ is a zero mean strictly stationary ergodic and strong mixing with mixing coefficients as in A(i) and $E|\eta_{rt}|^{2+\rho} < \infty$ for some $\rho > 0$, (ii) $S(q) = \sum_{\ell=0}^{\infty}|\gamma_\ell(q)| < M < \infty$.*

The above assumption C(i) ensures that an FCLT holds for the η_{rt} sequence as stated under our Assumption B. Specifically, $\sum_{t=1}^{[Ts]}\eta_{rt}/\sqrt{T} \Rightarrow B_{ur}(s)$. Note that in order for Assumption C to hold it suffices to require u_t to be strong mixing since under Assumption A(i), q_t and hence I_{rt-1} are strong mixing. The main motivation for the more primitive requirements stated in C comes from the need to use a martingale approximation for η_{rt} as described in Phillips (1988) and Hansen (1992). Part (ii) of Assumption C imposes a mild restriction on the joint dynamics of $\{u_t, v_t, q_t\}$ and is satisfied under the assumption that the series involved are strictly stationary and ergodic. We are now in a position to obtain a representation of the limiting distribution of our estimators.

Proposition 2. *Under assumptions A, B and C, as $T \to \infty$ and if $Th \to \infty$ and $Th^{3/2} \to 0$ we have*

$$
\begin{aligned}
\sqrt{Th}(\hat{f}_0(q) - f_0(q)) &\Rightarrow \frac{B_{ur}(1)}{E[I_{rt-1}/h]} \\
&\to \frac{1}{g_q(q)} N(0, \overline{\omega}^2) \quad as \quad h \to 0 \\
T\sqrt{h}(\hat{f}_1(q) - f_1(q)) &\Rightarrow \frac{\int_0^1 \overline{B}_v dB_{ur} + \sum_{\ell=0}^{\infty} E[\eta_{rt+\ell} v_t]}{E[I_{rt-1}/h] \int \overline{B}_v^2} \quad (7) \\
&\to MN(0, V(q)) \quad as \quad h \to 0 \quad (8)
\end{aligned}
$$

with $V(q) = \overline{\omega}^2/[g_q(q)]^2 \int \overline{B}_v^2$ and where $\overline{\omega}^2 = \lim_{h\to 0} E[u_t^2 I_{rt-1}/h]$.

The above asymptotics allow for very general dynamics across the shocks driving (1)-(2). It is useful to note that unlike the linear cointegration setting mixed normality arises despite the presence of endogeneity and possible serial correlation. At this stage it is also worth pointing out that if we impose $E[u_t^2|q_{t-1}] = E[u_t^2] \equiv \sigma_u^2$ then $\overline{\omega}^2 = \sigma_u^2 g_q(q)$ and we have

$$
\begin{aligned}
\sqrt{Th}(\hat{f}_0(q) - f_0(q)) &\Rightarrow N(0, \sigma_u^2) \\
T\sqrt{h}(\hat{f}_1(q) - f_1(q)) &\Rightarrow MN\left(0, \frac{\sigma_u^2}{g_q(q) \int \overline{B}_v^2}\right). \quad (9)
\end{aligned}
$$

4 Finite Sample Analysis

Our goal here is to illustrate the behaviour of our piecewise local linear estimators via a series of simulation experiments. We will consider five functional forms including one that violates our differentiability assumption in A(iii). The stochastic structure of our DGPs will be sufficiently general to allow for the presence of endogeneity and a rich dynamic structure for the errors driving x_t. Specifically, our DGP is given by

$$
\begin{aligned}
y_t &= f_0(q_{t-1}) + f_1(q_{t-1})\ x_t + u_t \\
x_t &= x_{t-1} + v_t \\
u_t &= \rho_u u_{t-1} + eu_t \\
v_t &= \rho_v v_{t-1} + ev_t \\
q_t &= \rho_q q_{t-1} + eq_{qt} \quad (10)
\end{aligned}
$$

and letting $z_t = (eu_t, ev_t, eq_t)'$ and $\mathbf{\Sigma}_z = E[z_t z_t']$, we use

$$\mathbf{\Sigma}_z = \begin{pmatrix} 1 & \sigma_{uv} & \sigma_{uq} \\ \sigma_{uv} & 1 & \sigma_{vq} \\ \sigma_{uq} & \sigma_{qv} & 1 \end{pmatrix}$$

for the covariance structure of the random disturbances. Note that our chosen covariance matrix parameterisation allows q_t to be contemporaneously correlated with the shocks to y_t and throughout all our experiments we set $\{\sigma_{uv}, \sigma_{uq}, \sigma_{vq}\} = \{-0.5, 0.5, 0.5\}$.

The range of possible functional coefficients we consider for either the intercept or the slope functions is given by

$$\begin{aligned} A: f(q) &= 0.3 - 0.5\ e^{-1.25q^2} \\ B: f(q) &= \frac{0.5}{1+e^{-4q}} - 0.75 \\ C: f(q) &= 0.25\ e^{-q^2} \\ D: f(q) &= 1 + 2(q > 0.5) \\ E: f(q) &= (1.5 + 0.6q)\ e^{-0.5(0.5q-1.5)^2} \end{aligned} \tag{11}$$

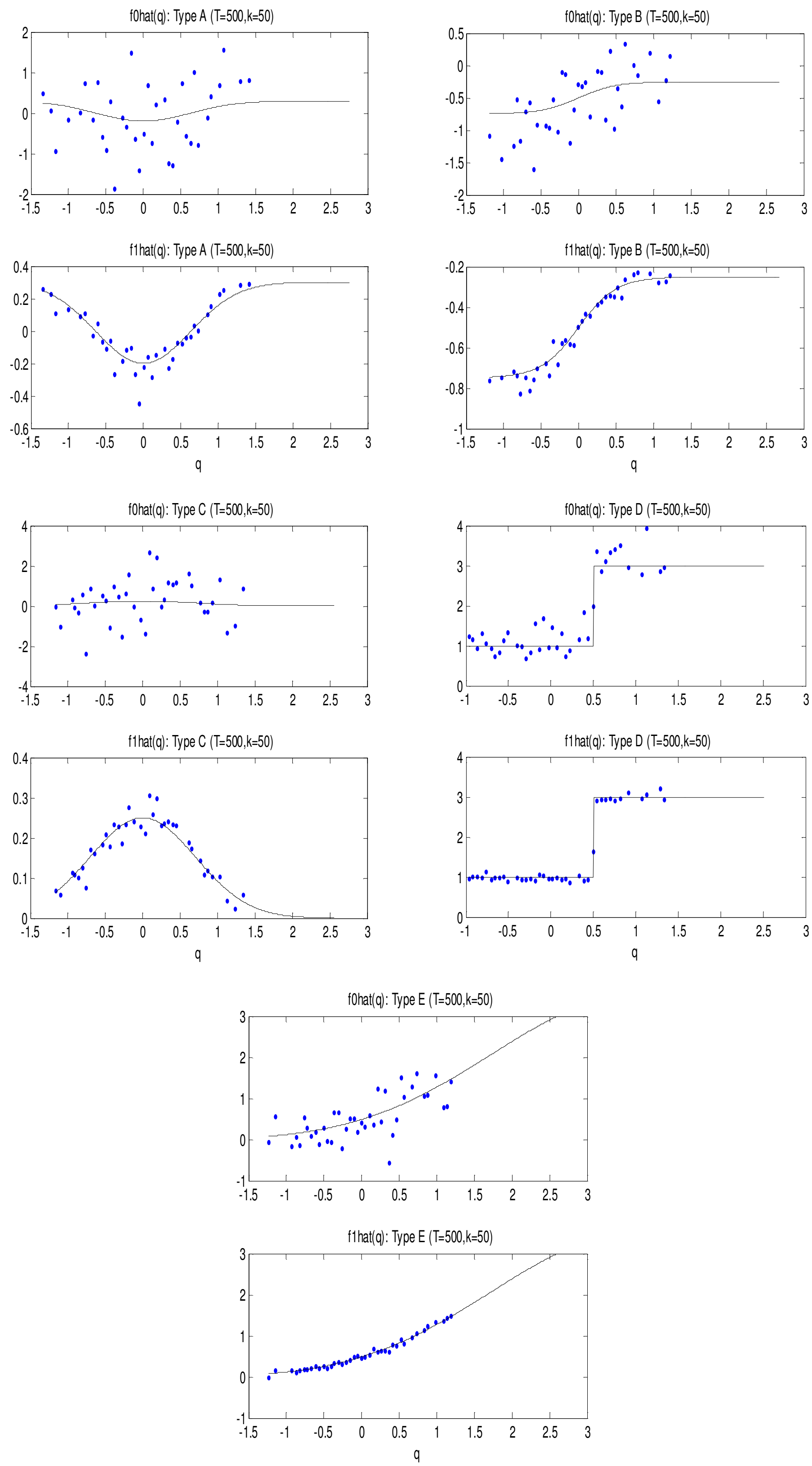
thus covering a very wide range of shapes including for illustration purposes a threshold type function which violates our differentiability assumption. Following standard practice in the functional coefficient literature, the quality of our estimators will be assessed via the computation of the root MSE defined as follows

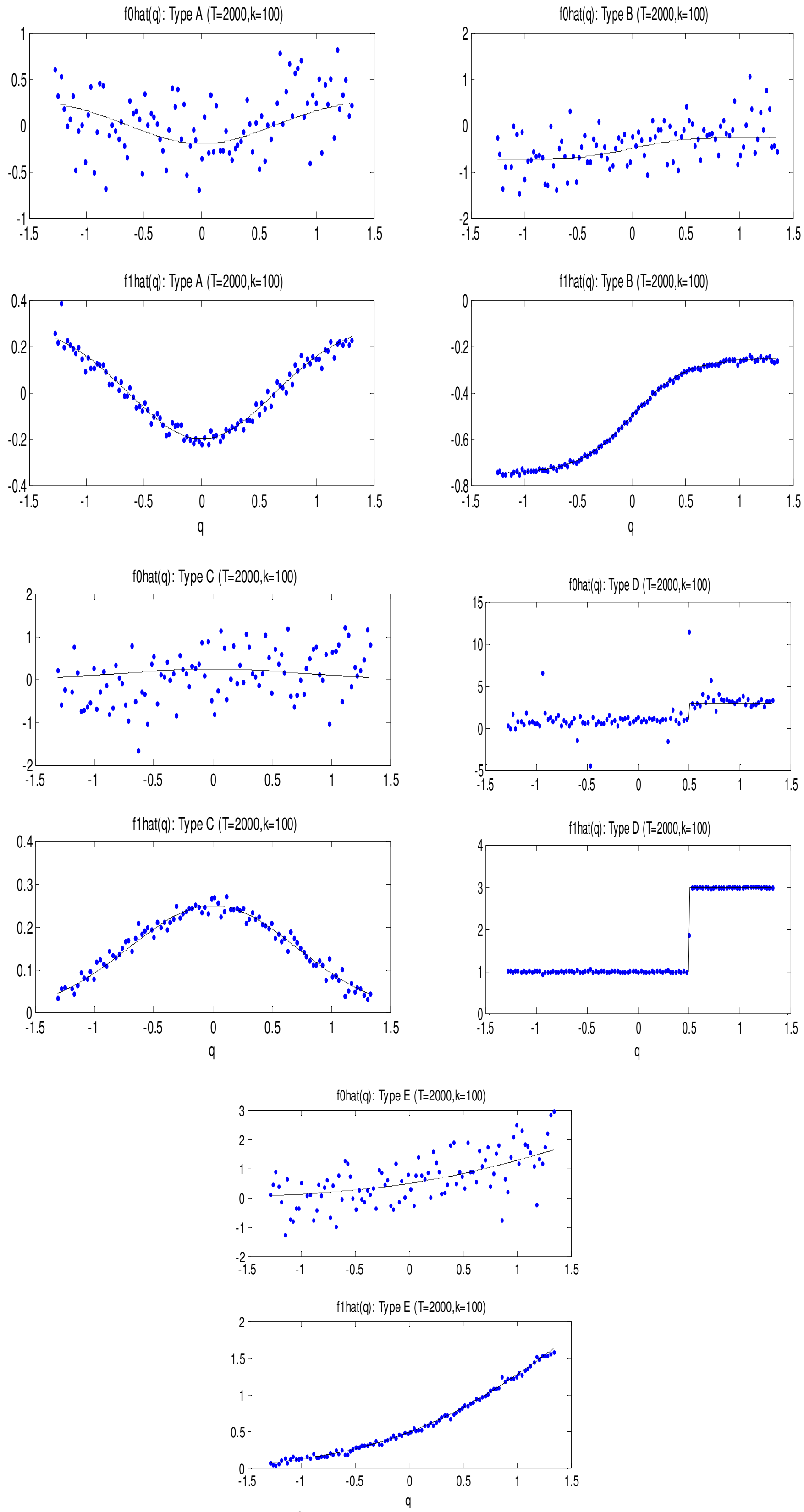
$$RMSE_i = \sqrt{\frac{1}{k}\sum_{r=1}^{k}(\hat{f}_i(q_r) - f_i(q_r))^2} \quad i = 0, 1 \tag{12}$$

for some q_r falling within each bin, say the midpoint. All our experiments use $NID(0,1)$ variables for the random disturbances z_t while setting $\{\rho_u, \rho_v, \rho_q\} = \{0.25, 0.25, 0.25\}$ thus allowing both serial correlation and endogeneity.

Before proceeding with our simulations we give a snapshot of the performance of our estimators by displaying plots of single realisation based $\hat{f}_i(q)'s$ for $i = 0, 1$ together with their true counterparts. Figure 1 below presents the plots of the functions corresponding to our formulations in A-E across samples of size T=500 and T=2000. The corresponding choice for the number of bins was k=50 and k=100.

Figure 1: Piecewise Local Linear Estimation under T=500 and T=2000


f0hat(q): Type A (T=500,k=50)
f0hat(q): Type B (T=500,k=50)
f1hat(q): Type A (T=500,k=50)
f1hat(q): Type B (T=500,k=50)
q
q
f0hat(q): Type C (T=500,k=50)
f0hat(q): Type D (T=500,k=50)
f1hat(q): Type C (T=500,k=50)
f1hat(q): Type D (T=500,k=50)
q
q
f0hat(q): Type E (T=500,k=50)
f1hat(q): Type E (T=500,k=50)
q


f0hat(q): Type A (T=2000,k=100)
f0hat(q): Type B (T=2000,k=100)
f1hat(q): Type A (T=2000,k=100)
f1hat(q): Type B (T=2000,k=100)
q
q
f0hat(q): Type C (T=2000,k=100)
f0hat(q): Type D (T=2000,k=100)
f1hat(q): Type C (T=2000,k=100)
f1hat(q): Type D (T=2000,k=100)
q
q
f0hat(q): Type E (T=2000,k=100)
f1hat(q): Type E (T=2000,k=100)
q

The above plots suggest that our $\hat{f}_1(q)$ estimator has an excellent ability to trace its true counterpart. Note also that these figures have been obtained allowing for both serial correlation and endogeneity in the underlying dynamics. It is also very interesting to note the performance of $\hat{f}_1(q)$ when the underlying true function has a kink as in scenario D. Both intercept and slope function estimators appear to match their true counterparts very accurately. However we must also emphasise the generally poor performance of $\hat{f}_0(q)$ when the sample size is small. This is clearly not unexpected and stems from the slow convergence of the estimator relative to that of $\hat{f}_1(q)$. Regardless of the sample size it is also evident that the variance of $\hat{f}_0(q)$ is substantially larger than that of $\hat{f}_1(q)$.

We next aim to highlight more formally the consistency properties of our estimators by documenting the progression of the corresponding RMSE's as the sample size is allowed to increase. Results across a selective set of scenarios are summarised in Table 1 below which displays simulated averages of (12) across N=2000 Monte-Carlo replications. The rows labelled *PLLE* correspond to our piecewise local linear estimator while the rows labelled *KERN* are based on a Kernel estimation as described in Xiao (2010) and using a Gaussian Kernel with $h = 1/k$ (the number of bins associated with each sample size is denoted k).

Table 1. RMSE of Estimators

		$T = 250$ $k = 25$	$T = 500$ $k = 50$	$T = 1000$ $k = 100$	$T = 2000$ $k = 200$	$T = 250$ $k = 25$	$T = 500$ $k = 50$	$T = 1000$ $k = 100$	$T = 2000$ $k = 200$
		$\hat{f}_0(q)$				$\hat{f}_1(q)$			
A	*PLLE*	0.6349	0.6201	0.5973	0.5551	0.0052	0.0025	0.0012	0.0006
	KERN	0.8416	2.0433	1.6833	2.9416	0.0047	0.0038	0.0027	0.0021
B	*PLLE*	0.6376	0.6204	0.6001	0.2249	0.0023	0.0011	0.0005	0.0000
	KERN	1.6611	119.1210	2.4946	0.0485	0.0037	1.2556	0.0015	0.0000
C	*PLLE*	0.6421	0.6116	0.5941	0.2308	0.0023	0.0011	0.0005	0.0000
	KERN	1.0415	4.8373	38.6487	0.0515	0.0055	0.0137	0.1126	0.0000
D	*PLLE*	6.0625	6.8851	5.4973	3.4962	0.0318	0.0173	0.0080	0.0022
	KERN	3.7597	3.8441	3.4285	1.7273	0.0748	0.0376	0.0183	0.0064
E	*PLLE*	0.6121	0.5831	0.5581	0.5517	0.0084	0.0041	0.0020	0.0010
	KERN	5.2090	8.0993	7.0420	29.2622	0.5476	0.1457	0.0647	0.0069

Across all functional forms we note a clear decline in the PLLE based RMSE's as the sample size is allowed to increase. As expected from Proposition 1, the slope functions see their RMSEs decline substantially faster than their intercept counterparts. Our comparisons with an alternative Kernel based estimator also suggest that our method competes favourably. Naturally, since alternative Kernels or functional forms may produce different finite sample outcomes it would be misleading to argue that our PLLE approach strictly dominates alternative approaches.

Our final goal is to explore the finite sample adequacy of the asymptotic normality result documented in our Proposition 2. It is important to note at this stage that the main goal of this paper is the estimation of the unknown functions $f_0(q)$ and $f_1(q)$ rather than the development of a full inference theory underlying these functions. For this reason we chose to concentrate solely on $\hat{f}_1(q)$ and empirically back up our statements in Proposition 2 through the simulation of the quantiles of the normalised quantity given by

$$\mathcal{Z}(q) = \frac{(\hat{f}_1(q) - f_1(q))}{\sqrt{\sigma_u^2 / \sum (x_t - \bar{x}_r)^2 I_{rt-1}}} \tag{13}$$

and which is distributed as N(0,1) under the scenario descibed in (10).

Our simulations set $\sigma_u^2 = 1$ and continue to allow for both endogeneity and serial correlation. We evaluate $\mathcal{Z}(q)$ at the midpoint, say q_r, of each bin $r = 1, \dots, k$ and obtain the relevant quantiles corresponding to each bin across our N replications i.e. k values of 2.5% cutoffs, k values of 5% cutoffs etc. Rather than displaying each of the resulting individual bin specific quantiles we average them across the k bins in order to obtain a compact metric of adequacy of the theoretical limiting distributions. Results are displayed in Table 2 below.

Table 2. Simulated Quantiles under T=1000 and k=100 (Averages across k bins)

	2.5%	5.0%	95.0%	97.5%
N(0, 1)	−1.96	−1.65	1.65	1.96
A	−2.07	−1.75	1.63	1.95
B	−2.07	−1.74	1.61	1.94
C	−2.05	−1.73	1.64	1.96
D	−3.40	−2.18	1.83	3.07
E	−2.08	−1.75	1.65	1.96

The above figures highlight a good to excellent match of the theory obtained under Proposition 2 with finite sample data across a wide range of functional forms. It is also interesting to note the inadequacy of the proposed theory for functions such as D which violate our differentiability assumptions.

5 Conclusions

This paper introduced the concept of functional cointegration and proposed a novel method of estimating the unknown functional coefficients linking the variables of interest under a nonstationary unit root setting. Our method is based on a simple binning idea and is shown to perform well asymptotically as well as in finite samples. Operating within a highly general probabilistic setting that allows for both serial correlation and endogeneity we established the consistency of our function estimators and also derived their limiting distribution. Since developing formal inferences was beyond the scope of this paper, in future work it will be interesting to use our results to obtain the properties of test statistics that could be used to tests hypotheses such as the null of a linearly cointegrated model versus our functional specification.

APPENDIX

LEMMA 1: As $T \to \infty$ and with $\overline{B}_v \equiv B_v - \int_0^1 B_v$ we have (i) $\sum I_{rt-1}/T \xrightarrow{p} E[I_{rt-1}]$, (ii) $\sum(q_{t-1} - q)I_{rt-1}/T \xrightarrow{p} E[(q_{t-1} - q)I_{rt-1}]$, (iii) $\sum u_t v_t I_{rt-1}/T \xrightarrow{p} E[u_t v_t I_{rt-1}]$, (iv) $\sum(x_t - \overline{x}_r)I_{rt-1}/T^{3/2} \Rightarrow E[I_{rt-1}]\int_0^1 \overline{B}_v$, (v) $\sum(x_t - \overline{x}_r)^2 I_{rt-1}/T^2 \Rightarrow E[I_{rt-1}]\int_0^1 \overline{B}_v^2$, (vi) $\sum(x_t - \overline{x}_r)(q_{t-1} - q)x_t I_{rt-1}/T^2 \Rightarrow E[(q_{t-1} - q)I_{rt-1}]\int_0^1 \overline{B}_v^2$ for some $q \in H_r$.

PROOFS: (i)-(iii) Assumption A ensures that I_{rt} is also strong mixing with the same mixing numbers as q_t. The results then follow from a suitable law of large numbers (see White (2001, Sections 3.3-3.4)). (iv) For notational simplicity we focus on $\sum x_t I_{rt}/T^{3/2}$ which we rewrite as $\sum x_t I_{rt}/T^{3/2} = \sum x_t(I_{rt} - p_r)/T^{3/2} + p_r \sum x_t/T^{3/2}$ with $p_r \equiv E[I_{rt}]$. Since $\sum x_t/T^{3/2} \Rightarrow \int B_v$ from Phillips (1987) it suffices to show that $\sum x_t(I_{rt} - p_r)/T^{3/2}$ is $o_p(1)$. Letting $S_t = \sum_{j=1}^t (I_{rj} - p_r)$ and using summation by parts we write $\sum_{t=1}^T x_t(I_{rt} - p_r)/T^{3/2} = (x_T/\sqrt{T})(S_T/T) - \sum_{t=1}^T S_{t-1}\Delta x_t/T^{3/2}$. Since $S_T/T \xrightarrow{p} 0$ and $x_T/\sqrt{T} = O_p(1)$ it follows that $x_T S_T/T\sqrt{T} \xrightarrow{p} 0$. Similarly, $\sum S_{t-1}\Delta x_t = \sum v_t S_{t-1}$ is such that $E[\sum v_t S_{t-1}]^2 = O(T^2)$ implying that $\sum v_t S_{t-1} = O_p(T)$ and thus ensuring that $\sum v_t S_{t-1}/T^{3/2} = o_p(1)$ as required. The proofs of (v)-(vi) follow identical lines and are therefore omitted.

LEMMA 2: As $h \to 0$ (i) $E[I_{rt-1}]/h \to g_q(q)$, (ii) $E[I_{rt-1}(q_{t-1} - q)^m] = o(h^{m+1})$.

PROOF: We focus on (ii) and evaluate the expression at some $q = q_r$. We have

$$
\begin{aligned}
|E[(q_{t-1} - q_r)^m I_{rt-1}]| &= \left| \int_{H_r} (q - q_r)^m g_q(q) dq \right| \\
&\leq \int_{H_r} |q - q_r|^m g_q(q) dq \\
&\leq h^m \int_{H_r} g_q(q) dq = const * \ h^{m+1}
\end{aligned} \tag{14}
$$

and the result follows.

PROOF OF PROPOSITION 1: Given x_t, y_t, q_t and the known bin cutoff locations the least squares estimators of the intercept β_{0r} and slope parameter β_{1r} of the regression line within each bin can be formulated as

$$
\begin{aligned}
\hat{\beta}_{0r} &= \overline{y}_r - \hat{\beta}_{1r}\overline{x}_r \\
\hat{\beta}_{1r} &= \frac{\sum(x_t - \overline{x}_r)I_{rt-1}y_t}{\sum(x_t - \overline{x}_r)^2 I_{rt-1}}
\end{aligned} \tag{15}
$$

with $\overline{x}_r = \sum x_t I_{rt-1}/\sum I_{rt-1}$ and $\overline{y}_r = \sum y_t I_{rt-1}/\sum I_{rt-1}$. Next, using $y_t = f_0(q_{t-1}) + f_1(q_{t-1})x_t + u_t$, taking a first order Taylor expansion of the unknown coefficients around some $q \in H_r$

$$
f_i(q_{t-1}) \approx f_i(q) + f_i'(q)(q_{t-1} - q) + o(h^2) \tag{16}
$$

for $i = 0, 1$ and ignoring terms that are $o(h^2)$ we can rewrite $\hat{\beta}_{1r}$ as

$$\begin{aligned}\hat{\beta}_{1r} - f_1(q) &= \frac{\sum(x_t - \overline{x}_r)I_{rt-1}[f_0(q_{t-1}) + f_1(q_{t-1})x_t]}{\sum(x_t - \overline{x}_r)^2 I_{rt-1}} + \frac{\sum(x_t - \overline{x}_r)I_{rt-1}u_t}{\sum(x_t - \overline{x}_r)^2 I_{rt-1}} \\ &= f_0'(q)\frac{\sum(x_t - \overline{x}_r)(q_{t-1} - q)I_{rt-1}}{\sum(x_t - \overline{x}_r)^2 I_{rt-1}} + f_1'(q)\frac{\sum x_t(x_t - \overline{x}_r)(q_{t-1} - q)I_{rt-1}}{\sum(x_t - \overline{x}_r)^2 I_{rt-1}} \\ &+ \frac{\sum(x_t - \overline{x}_r)I_{rt-1}u_t}{\sum(x_t - \overline{x}_r)^2 I_{rt-1}}. \end{aligned} \tag{17}$$

It is now also convenient to reformulate the above as

$$\begin{aligned} T\sqrt{h}(\hat{\beta}_{1r} - f_1(q)) &= f_0'(q)\left(\frac{\sum(x_t - \overline{x}_r)(q_{t-1} - q)I_{rt-1}/T^2h}{\sum(x_t - \overline{x}_r)^2 I_{rt-1}/T^2h}\right)T\sqrt{h} + \\ &\quad f_1'(q)\left(\frac{\sum x_t(x_t - \overline{x}_r)(q_{t-1} - q)I_{rt-1}/T^2h}{\sum(x_t - \overline{x}_r)^2 I_{rt-1}/T^2h}\right)T\sqrt{h} + \\ &\quad \frac{\sum(x_t - \overline{x}_r)I_{rt-1}u_t/T\sqrt{h}}{\sum(x_t - \overline{x}_r)^2 I_{rt-1}/T^2h} \\ &\equiv T\sqrt{h}\ A_{rt} + T\sqrt{h}\ B_{rt} + C_{rt} \end{aligned} \tag{18}$$

and the result follows by showing that $T\sqrt{h}\ A_{rt}$ and $T\sqrt{h}\ B_{rt}$ are asymptotically negligible when $Th^{3/2} \to 0$ while C_{rt} is $O_p(1)$. From Lemma 2(i)-(ii) we have

$$B_{rt} \Rightarrow f_1'(q)\frac{E[(q_{t-1} - q)I_{rt-1}/h]}{E[I_{rt-1}/h]} \equiv B_{1\infty} \tag{19}$$

and using Lemma 2 it is also clear that $B_{1\infty} = o(h)$ as needed since we operate under $Th^{3/2} \to 0$. The asymptotic negligibility of $T\sqrt{h}\ A_{rt}$ follows along identical lines using Lemma 2. Finally, for C_{rt} we concentrate solely on its numerator since from Lemma 1(v) its denominator is known to be $O_p(1)$ since converging in distribution. Using $x_t = x_{t-1} + v_t$ we write

$$\frac{\sum(x_t - \overline{x}_r)I_{rt-1}u_t}{T\sqrt{h}} = \frac{\sum x_{t-1}I_{rt-1}u_t}{T\sqrt{h}} + \frac{\sum u_t v_t I_{rt-1}}{T\sqrt{h}} - \left(\frac{\sum x_t I_{rt-1}/T\sqrt{T}}{\sum I_{rt-1}/T}\right)\frac{\sum u_t I_{rt-1}}{\sqrt{Th}}. \tag{20}$$

From Lemma 1(iii) $\sum u_t v_t I_{rt-1}/T \stackrel{p}{\to} E[u_t v_t I_{rt-1}]$ so that the second term in the right hand side of (20) is $o(\sqrt{h})$ while the third term is $O_p(1)$ by Lemma 1 and Assumption B so that we can concentrate on $\sum x_{t-1}u_t I_{rt-1}/T\sqrt{h}$. We write

$$\begin{aligned}\left|\frac{1}{T\sqrt{h}}\sum x_{t-1}u_t I_{rt-1}\right| &\leq \max_{t\leq T}\left|\frac{x_t}{\sqrt{T}}\right|\frac{1}{\sqrt{Th}}\sum |u_t|I_{rt-1} \\ &\leq \left(\sup_{s\in[0,1]} B_v(s) + 1\right)\frac{1}{\sqrt{Th}}\sum |u_t|I_{rt-1} = O_p(1) \end{aligned} \tag{21}$$

and hence leading to the required result.

Proceeding along the same lines for $\hat{\beta}_{0r}$ and using $\hat{\beta}_{1r} = f_1(q) + O_p(1/T\sqrt{h})$ we write

$$\begin{aligned}\hat{\beta}_{0r} - f_0(q) &= f_0'(q)\frac{\sum(q_{t-1} - q)I_{rt-1}}{\sum I_{rt-1}} + f_1'(q)\frac{\sum(q_{t-1} - q)x_t I_{rt-1}}{\sum I_{rt-1}} + \\ &\quad \frac{\sum u_t I_{rt-1}}{\sum I_{rt-1}} - \overline{x}_r O_p(\frac{1}{T\sqrt{h}}). \end{aligned} \tag{22}$$

Applying suitable normalisations we reformulate (22) as

$$\sqrt{Th}(\hat{\beta}_{0r} - f_0(q)) = f_0'(q)\left(\frac{\sum(q_{t-1}-q)I_{rt-1}}{\sum I_{rt-1}}\right)\sqrt{Th} + \left(f_1'(q)\frac{\sum(q_{t-1}-q)x_t I_{rt-1}}{\sum I_{rt-1}}\right)\sqrt{Th} + \frac{\sum u_t I_{rt-1}/\sqrt{Th}}{\sum I_{rt-1}/Th} + O_p(1/\sqrt{Th}). \quad (23)$$

Proceeding as above and using Lemmas 1-2 it is again straightforward to observe that under $\sqrt{T}h^{3/2} \to 0$ the first two terms in the right hand side of (23) are asymptotically negligible while the third term is $O_p(1)$ by our Assumptions A-B.

PROOF OF PROPOSITION 2: From our Assumption A we have $\sum u_t I_{rt-1}/\sqrt{Th} \Rightarrow B_{ur}(r)$ while from Lemma 1 (i) we have $\sum I_{rt-1}/T \stackrel{p}{\to} E[I_{rt-1}]$. The result in (7) then follows directly from the continuous mapping theorem, noting also that $E[I_{rt-1}/h] \to g_q(q)$ as $h \to 0$. Regarding (8), since we operate under $Th^{3/2} \to 0$, from (19) it follows that we can concentrate on the limiting behaviour of $C_{rt} = \sum((x_t - \bar{x}_r)u_t I_{rt-1}/T\sqrt{h})/\sum(x_t - \bar{x}_r)^2 I_{rt-1}/T^2 h$. Given our result in Lemma 1(v) we consider solely the numerator of C_{rt}, and for simplicity we can focus on the simpler quantity $\sum x_{t-1}\eta_{rt}/T\sqrt{h} \equiv NC_{rt}$ with $\eta_{rt} = u_t I_{rt-1}$. Letting $\mathcal{F}_t = \sigma(\eta_{rt}, \eta_{rt-1}, \ldots, v_t, v_{t-1}, \ldots)$ we follow Phillips (1988) and Hansen (1992) and make use of the martingale approximation $\eta_{rt} = e_{rt} + z_{rt-1} - z_{rt}$ where $e_{rt} = \sum_{\ell=0}^{\infty}(E[\eta_{rt+\ell}|\mathcal{F}_t] - E[\eta_{rt+\ell}|\mathcal{F}_{t-1}])$ and $z_{rt} = \sum_{\ell=1}^{\infty} E[\eta_{rt+\ell}|\mathcal{F}_t]$. We can now write $\sum x_{t-1}\eta_{rt} = \sum x_{t-1}e_{rt} - \sum(z_{rt} - z_{rt-1})x_{t-1}$ and equivalently

$$\frac{1}{T\sqrt{h}}\sum x_{t-1}\eta_{rt} = \frac{1}{T\sqrt{h}}\sum x_{t-1}e_t + \frac{1}{T\sqrt{h}}\sum v_t z_t - \frac{R_{rT}}{\sqrt{h}} \quad (24)$$

with $R_{rT} = z_{rT}x_T/T$ and $\Delta x_t = v_t$. Clearly $\max_t |z_{rt}x_t/T| \leq \left|\sum_{j=1}^{T} v_j/\sqrt{T}\right| \max_t |z_{rt}|/T \stackrel{p}{\to} 0$ since $\max_t |z_{rt}|/T \stackrel{p}{\to} 0$ from Hansen (1992). We can therefore write

$$\frac{1}{T\sqrt{h}}\sum x_{t-1}\eta_{rt} = \frac{1}{T\sqrt{h}}\sum x_{t-1}e_t + \frac{1}{T\sqrt{h}}\sum v_t z_t + o_p(1). \quad (25)$$

Since $\{e_t, \mathcal{F}_t\}$ is a martingale difference sequence and $x_{[Tr]}/\sqrt{T} \Rightarrow B_v(r)$ by our Assumption B it follows from Kurtz and Protter (1991) and Hansen (1992) that $\sum x_{t-1}e_t/T\sqrt{h} \Rightarrow \int B_v dB_{ur}$. The result in (7) then follows by noting that $E[\sum_{\ell=1}^{\infty} E[\eta_{rt+\ell}v_t|\mathcal{F}_t]] = \sum_{\ell=1}^{\infty} E[v_t\eta_{rt+\ell}]$. More specifically, using (21) and the continuous mapping theorem gives

$$\begin{aligned}\frac{\sum(x_t - \overline{x}_r)I_{rt-1}u_t}{T\sqrt{h}} &\Rightarrow \int_0^1 B_v dB_{ur} + \sum_{\ell=1}^{\infty} E[\eta_{rt+\ell}v_t] + E[\eta_{rt}v_t] - B_{ur}(1)\int_0^1 B_v \\ &= \int \overline{B}_v dB_{ur} + \sum_{\ell=0}^{\infty} E[\eta_{rt+\ell}v_t] \quad (26)\end{aligned}$$

as required. The result in (9) follows noting that $E[\eta_{rt+\ell} v_t] = \int_{H_r} \gamma_0(q)/\sqrt{h} = o(\sqrt{h})$ and

$$\begin{aligned}
\left| \frac{1}{\sqrt{h}} \sum_{\ell=1}^{\infty} E[u_{t+\ell} I_{rt+\ell-1} v_t] \right| &= \left| \frac{1}{\sqrt{h}} \sum_{\ell=1}^{\infty} E[I_{rt+\ell-1} E[u_{t+\ell} v_t | q_{t+\ell-1}]] \right| \\
&= \frac{1}{\sqrt{h}} \left| \sum_{\ell=1}^{\infty} \int_{H_r} \gamma_\ell(q) g(q) dq \right| \\
&\leq \frac{1}{\sqrt{h}} \int_{H_r} \sum_{\ell=1}^{\infty} |\gamma_\ell(q)| g(q) dq \\
&\leq \frac{1}{\sqrt{h}} M \int_{H_r} g(q) dq \\
&= o(\sqrt{h}) \qquad (27)
\end{aligned}$$

which follows directly from our Assumption C(ii).

APPENDIX

Banerjee, A. N., 1994. A method of estimating the average derivative. CORE Discussion Paper: 9403, Université Catholique de Louvain.

Banerjee, A. N., 2007. A method of estimating the average derivative. Journal of Econometrics, 136, 65-88.

Berenguer, V., 2010. Summability of Stochastic Processes: A Generalization of Integration and Co-integration valid for Non-linear Processes, Universidad Carlos III de Madrid Working Paper.

Berenguer, V. and Gonzalo, J. 2011. Co-Summability: From Linear to Non-Linear Co-Integration, Universidad Carlos III de Madrid Working Paper.

Breitung, J., 2001. Rank tests for nonlinear cointegration, Journal of Business and Economic Statistics, 19, 331-340.

Cai, Z., Li, Q. and Park, J. Y., 2009. Functional Coefficient Models for Nonstationary Time Series Data, Journal of Econometrics, 148, 101-113.

Cleveland, W.S., Grosse, E., and Shyu, W.M., 1991. Local Regression Models. In *Statistical Models in S* (J.M. Chambers and T.J. Hastie, ed.). Wadsworth & Brooks, Pacific Grove. CA. pp. 309-376.

Chen, R. and Tsay, R.S. 1993. Functional Coefficient Autoregressive Models, Journal of the American Statistical Association, 88, 298-308.

Fan, J. and Zhang, W. 1999. Statistical Estimation in Varying Coefficient Models, The Annals of Statistics, 27, 1491-1518.

Fan, J. and Yao, Q. 2003. Nonlinear Time Series. Springer-Verlag, New-York.

Fan, J. and Gijbels, I. 1992. Variable Bandwidth and Local Linear Regression Smoothers, The Annals of Statistics, 20, 2008-2036.

Fan, J. and Gijbels, I. 1996. Local Polynomial Modelling and its Applications. Chapman and Hall, London.

Gonzalo, J. and Pitarakis, J., 2006. Threshold Effects in Cointegrating Relationships, Oxford Bulletin of Economics and Statistics, 68, 813-833.

Granger, C.W.J. and Hallman, J. 1991. Nonlinear Transformations of Integrated Time Series, Journal of Time Series Analysis, 12, 207-234.

Hansen, B., 1992. Convergence to Stochastic Integrals for Dependent Heterogeneous Processes, Econometric Theory, 8, 489-500.

Hastie, T. J. and Tibshirani, R. J., 1993. Varying Coefficient Models, Journal of the Royal Statistical Society, Series B, 55, 757-796.

Juhl, T., 2005. Functional-Coefficient Models under Unit Root Behaviour, Econometrics Journal, 8, 197-213.

Karlsen, H. A., Myklebust, T. and Tjostheim, D. 2007. Nonparametric estimation in a nonlinear cointegration type model, Annals of Statistics, 35, 252-299.

Kasparis, I. and Phillips, P. C. B., 2009. Dynamic Misspecification in Nonparametric Cointegrating Regression. Yale University Working Paper

Kurtz, T.G. and Protter, P., 1991. Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations, The Annals of Probability, 19, 1035-1070.

Phillips, P.C.B. 1987. Time series regression with unit roots, Econometrica, 55, 277-302.

Saikkonen, P. and Choi, I., 2004. Testing linearity in cointegrating smooth transition regressions, Econometrics Journal, 7, 341-365.

Senturk, D. and Mueller, H. G. (2005). Covariate-adjusted regression. Biometrika, 92, 75-89.

Senturk, D. and Mueller, H. G. (2006). Inference for covariate adjusted regression via varying coefficient models. Annals of Statistics, 34, 654-679.

Wang, Q. and Phillips, P.C.B. 2009. Structural Nonparametric Cointegrating Regression, Econometrica, 77, 1901-1948.

White, H., 2001. Asymptotic Theory for Econometricians. Academic Press.

Xiao, Z., 2009. Functional-Coefficient Cointegration Models, Journal of Econometrics, 152, 81-92.