DE GRUYTER DOI 10.1515/snde-2013-0083 — Stud. Nonlinear Dyn. E. 2014; 18(5): 507–520

Anurag Banerjee and Jean-Yves Pitarakis*

Functional cointegration: definition and nonparametric estimation

Abstract: We formally define a concept of functional cointegration linking the dynamics of two time series via a functional coefficient. This is achieved through the use of a concept of summability as an alternative to I(1)'ness which is no longer suitable under nonlinear dynamics. We subsequently introduce a nonparametric approach for estimating the unknown functional coefficients. Our method is based on a piecewise local least squares principle and is computationally simple to implement. We establish its consistency properties and evaluate its performance in finite samples. We subsequently illustrate its usefulness through an application that explores linkages between stock prices and dividends via a sentiment indicator.

Keywords: cointegration; functional coefficients; piecewise local linear estimation; unit roots.

JEL: C22; C50.

***Corresponding author: Jean-Yves Pitarakis,** University of Southampton, School of Social Sciences, Division of Economics, Highfield, Southampton, SO17 1BJ, UK, e-mail: j.pitarakis@soton.ac.uk
Anurag Banerjee: Durham University Business School, Mill Hill Lane, Durham, DH1 3LB, UK

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.

Our initial objective in this paper is to formally define a novel concept of functional cointegration linking two highly persistent variables via functional coefficients. Our framework is analogous to the well known linear cointegration property linking I(1) variables except that in the present nonlinear framework I(1)'ness is no longer suitable for describing the stochastic properties of our variables. 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 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. Our subsequent 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 our nonstandard context that allows for the presence of a unit root variable as in the recent work of Juhl (2005), Xiao (2009) and Cai, Li and Park (2009). Our method is very different from kernel smoothing based methods, relies on a straightforward disjoint binning principle and does not generally require the differentiability of the density of q and is shown to have good finite sample properties. An additional convenient aspect of our approach is its reliance on standard model selection based procedures (e.g. AIC, BIC) for the determination of the number of bins. This is akin to choosing a suitable bandwidth in kernel based methods.

The plan of the paper is as follows. Section 2 introduces and motivates our model and formally defines the concept of functional cointegration. Section 3 describes our estimation methodology and derives its asymptotic properties. Section 4 explores its performance in finite samples. Section 5 illustrates our estimation methodolody with an example linking stock prices and dividends via a sentiment proxy and Section 6 concludes. All proofs are relegated to the appendix.

2 The model and motivations

We consider the following functional coefficient regression model

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

$$x_t=x_{t-1}+v_t \tag{2}$$

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 smoother variants such as $f_i(q_{t-1})=[1+e^{-\gamma_i(q_{t-1}-c_i)}]^{-1}$ amongst others.

Before proceeding with the estimation of 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 $y_t=f_tx_t+u_t$ 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-Rico (2010) and Berenguer-Rico 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 obtaining reliable estimates of the unknown functional coefficients $f_0(q)$ and $f_1(q)$ and exploring their consistency properties. For this purpose we introduce a piecewise linear estimation approach initially 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. We partition the support of q_{t-1} into k disjoint bins of equal length $|H_r|=h$, $r=1,\dots,k$ (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 $S_{xx}^{(r)}=\sum_{t=1}^{T}\tilde{x}_t\tilde{x}_t'I_{rt-1}$ and $S_{xy}^{(r)}=\sum_{t=1}^{T}\tilde{x}_t y_t I_{rt-1}$ with $I_{rt-1}\equiv I_r(q_{t-1})$. Note that $\hat{\beta}_r$ provides the least-squares estimators of the intercept and slope parameters characterising the linear regression line within each bin. 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 $I_r(q)=I(q\in H_r)$.

At this stage it is important to emphasise that our method is fundamentally different from kernel smoothing based approaches (e.g., local linear regression with a uniform or any other type of kernel) since it relies on the use of disjoint bins. Local linear regression with a uniform kernel for instance is a weighted least squares estimator with the equal weights centered around the q observations [see Fan and Zhang (1999)]. Estimating the function at two different locations, say q_1 and q_2, via this approach will involve the use of some common data points (overlapping data) whereas under PLLE the estimates obtained at two different bins will not use any common data points since the bins are disjoint. 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.

Having introduced the mechanics behind our estimator our main goal is to establish its consistency. 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 assumptions. The following baseline assumptions will be maintained throughout the entire paper where we let $q_t=\mu+u_{qt}$.

Assumptions A. (i) $w_t=\{u_t, v_t, u_{qt}\}$ *is such that* $E[w_t]=0$, $E\|w_0\|^{\rho+\epsilon}<\infty$ *for some* $\rho>2$ *and the sequence* $\{w_t\}$ *is strictly stationary, strong mixing with mixing coefficients* α_m *such that* $\sum\alpha_m^{1-2/\rho}<\infty$. *(ii) The density of q denoted* $g_q(q)$ *is strictly positive and satisfies* $\sup_q g_q(q)<\underline{c}<\infty$ *and* $\inf_q g_q(q)>\underline{c}>0$. *(iii)* $g_q(q)$ *has compact support. (iv) The functional coefficients are twice continuously differentiable in q.*

Assumptions A above impose a very standard set of restrictions on the dynamics driving (1)–(2) leaving all random disturbances to be flexible enough to display rich dynamics such as ARMA process. Their joint interactions is also allowed to be very flexible since u_t and v_t can be correlated at all leads and lags and similarly for the interactions between q_t and the remaining variables included in the model. It is naturally understood that the associated long run variances of those processes are positive. In this sense the above setting is at least as flexible as the well known linear cointegration model formulated in triangular form allowing for both serial correlation and endogeneity. Note also that the strictly stationary and strong mixing nature of u_{qt} also implies that the indicator function series I_{rt} are strictly stationary and strong mixing with the same mixing coefficients.

Assumption A(ii) is concerned with the density of q_t and is required so as to ensure that there are observations in each bin. Since our estimation methodology requires fitting a least squares line within each bin of length $|H_r|=h$ it is understood throughout this paper that for estimability purposes there are enough observations falling within each bin. Note, however, that we do not impose any smoothness conditions on the density of q. This is in contrast with other methods that have been used in the literature (e.g., kernel smoothing via local linear regression). Assumption A(iii) requires the support of q to be compact. More specifically we require q to be bounded from below and above. In practice and throughout our simulations we form the support of q_t by taking the range of a top (say 0.9) and bottom (say 0.1) quantile. Finally, the differentiability of the $f_i(q)$*'s* will allow us to use their local Taylor expansions at a point q within each bin.

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, 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 has focused on the consistency of our proposed estimators 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 least squares based slope estimator converges at rate T while the intercept parameter estimator 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 2005; Xiao 2009).

4 Finite sample analysis

Our goal here is to illustrate the behaviour of our piecewise local linear estimators via a series of simulation experiments and also compare their performance with alternative kernel based estimation approaches.

We consider five functional forms including one that violates our differentiability assumption in A(iv). The stochastic structure of our DGPs is 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

$$\begin{aligned} y_t &= f_0(q_{t-1}) + f_1(q_{t-1})\, x_t + u_t \\ x_t &= x_{t-1} + v_t \\ u_t &= \rho_u u_{t-1} + eu_t \\ v_t &= \rho_v v_{t-1} + ev_t \\ q_t &= \rho_q q_{t-1} + eq_{qt}. \end{aligned} \tag{5}$$

Letting $z_t=(eu_t, ev_t, eq_t)'$ and $\Sigma_z = E[z_t z_t']$, we use

$$\Sigma_z = \begin{pmatrix} 1 & \sigma_{uv} & \sigma_{uq} \\ \sigma_{uv} & 1 & \sigma_{vq} \\ \sigma_{uq} & \sigma_{qv} & 1 \end{pmatrix}$$

for the covariance structure of the random disturbances. 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

$$\begin{aligned} A&: f(q) = \frac{200}{1+e^{-0.65q}} - 10 \\ B&: f(q) = 0.3 - 0.5\, e^{-1.25q^2} \\ C&: f(q) = 0.25\, e^{-q^2} \\ D&: f(q) = 1 + 2(q>0.5) \\ E&: f(q) = (1.5+0.6q)\, e^{-0.5(0.5q-1.5)^2} \\ F&: f(q) = \frac{e^{40q}}{1+e^{40q}} \end{aligned} \tag{6}$$

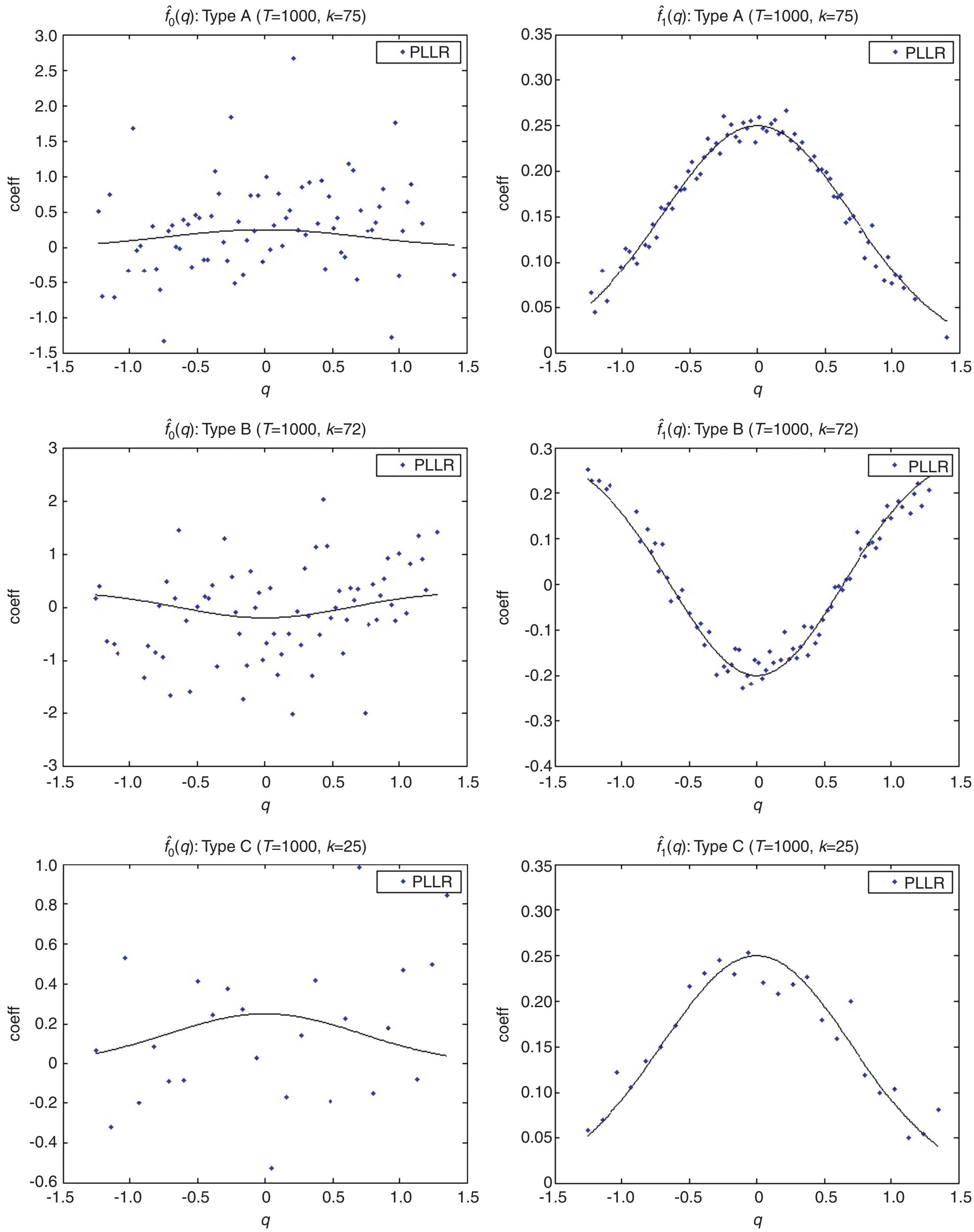
thus covering a very wide variety of shapes including for illustration purposes a threshold type function which violates our differentiability assumption. Following standard practice in the functional coefficient literature, the quality of our estimators will be assessed via the computation of the root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

$$RMSE_i = \sqrt{\frac{1}{k}\sum_{r=1}^{k}(\hat{f}_i(q_r) - f_i(q_r))^2} \quad i=0,1 \tag{7}$$

for some q_r falling within each bin, say the midpoint (note that since we operate under piecewise linearity the location at which we evaluate the function within the bin does not affect its value). 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 Figure 1 gives a snapshot of the performance of our estimators by displaying the plots of single realisation based $\hat{f}_i(q)'s$ for $i=0, 1$ together with their true counterparts. The number of bins associated with each example has been chosen via a formal model selection procedure discussed in greater detail below.

The plots suggest that $\hat{f}_1(q)$ displays a good ability to trace its true counterpart $f_1(q)$ along the chosen domain. Interestingly, our estimator also appears to match its true counterpart closely under scenario D when the chosen functional form has a kink. At this stage it is also worth recalling that these figures have been obtained allowing for both serial correlation and endogeneity in the underlying dynamics. Unlike $\hat{f}_1(q)$



(**Figure 1** Continued)

however, the estimator of $f_0(q)$ appears to perform poorly overall especially when the sample size is small. This is not unexpected and stems from the slow convergence of the estimator relative to that of $\hat{f}_1(q)$ as it occurs in the linear cointegration framework. It is also clear from the above plots 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 RMSEs as the sample size and associated bin number is allowed to increase.

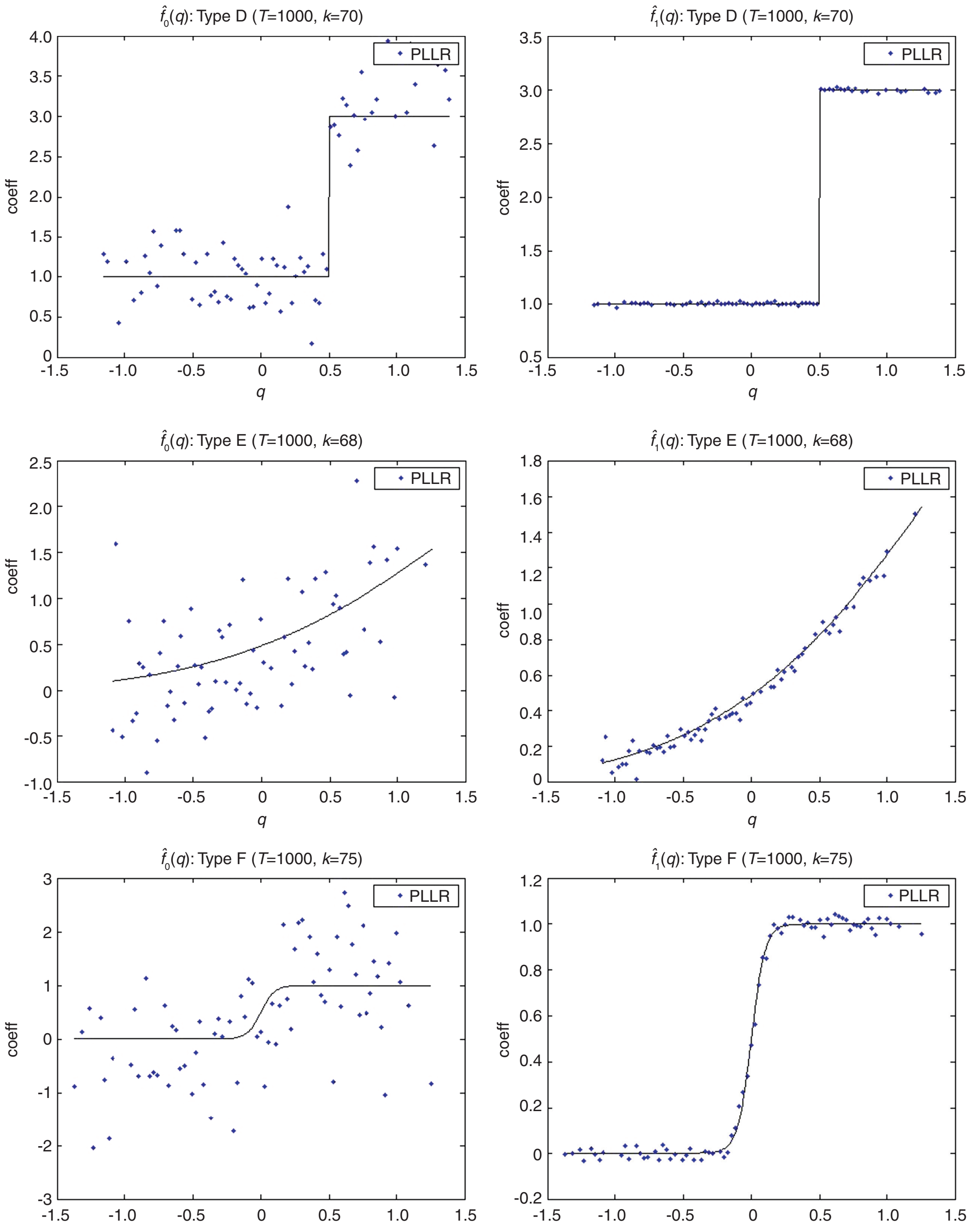


Figure 1 Piecewise local linear estimation.

Before proceeding further however it is important to recall that the implementation of our piecewise local linear approach requires the user to input the number of disjoint bins for partitioning the support of q_{t-1}. This user input is in a way similar to the bandwidth choice requirement that arises under kernel based estimation methods.

Since our estimation problem in (4) can be rewritten as a least squares regression problem with dummy variables associated with each bin and shifting both intercept and slope parameters, we propose to select an optimal k via a standard model selection based approach. We view this as an additional advantage of our approach. Note that the estimator in (4) can be viewed as equivalent to estimating the parametric model $y_t = \sum_{r=1}^{k} \beta_{0k} I_r(q_t) + \sum_{r=1}^{k} \beta_{1r} I_r(q_t) x_t + u_t$ with $2k$ parameters. Starting with a lower bound k_{min} for the number of bins and an upperbound k_{max} we select the optimal number of bins via the minimisation of an AIC type of criterion. More specifically, we implement our Piecewise Local Linear estimation method for each possible value of $k=k_{min}, \ldots, k_{max}$ and select the optimal number of bins as the one that leads to the smallest AIC value. As discussed in our simulations below we have also experimented with alternative criteria such as the BIC and $\bar{R}^2$ with almost identical outcomes to the ones obtained under the AIC and virtually no changes in the resulting RMSEs.

For comparison purposes we also contrast the performance of our method with Xiao's kernel based local linear approach as discussed in Xiao (2009). The latter is implemented using the particular fourth order kernel given by $k(u)=15(7u^4-10u^2+1)I(|u|\leq 1)/32$ and bandwidth $h=s_q/T^{1/3}$ with s_q denoting the standard deviation of q as advocated in the finite sample performance study of Xiao (2009). In order to highlight and emphasise the fundamental differences between our piecewise local linear model that uses disjoint bins from a kernel based approach with a uniform kernel, our simulations also consider Xiao's estimation approach with the use of a uniform kernel $k(u)=0.5I(|u|\leq 1)$.

Results across a selective set of scenarios are summarised in Table 1 below which displays simulated averages of (7) across N=2000 Monte-Carlo replications. The rows labelled *PLLE* correspond to our piecewise local linear estimator while the rows labelled *KER1* and *KER2* are based on a Kernel estimation as described within the finite sample performance analysis section of Xiao (2009) using the quadratic and uniform kernels respectively. In order to maintain comparability, the kernel based RMSEs in (7) have been evaluated at the same q_r*'s* as the ones dictated by our optimal number of bins.

Table 1 RMSE of Estimators under Serial Correlation and Endogeneity.

	T=250	T=500	T=1000	T=250	T=500	T=1000
	k_{max}=40	k_{max}=80	k_{max}=120	k_{max}=40	k_{max}=80	k_{max}=120
			$\hat{f}_0(q)$			$\hat{f}_1(q)$
A						
PLLE	8.950	6.292	4.935	0.573	0.286	0.174
KER1	45.858	12.230	8.843	6.162	0.704	0.382
KER2	11.710	10.289	6.686	1.119	0.720	0.338
B						
PLLE	0.981	0.904	0.760	0.085	0.056	0.025
KER1	1.535	0.584	2.245	0.308	0.174	0.040
KER2	0.395	0.319	0.267	0.036	0.020	0.011
C						
PLLE	0.875	0.885	0.812	0.081	0.056	0.036
KER1	5.817	0.526	0.396	0.189	0.033	0.020
KER2	0.373	0.298	0.248	0.034	0.019	0.011
D						
PLLE	1.405	1.170	0.942	0.111	0.071	0.042
KER1	14.728	4.542	4.346	0.353	0.191	0.272
KER2	1.272	1.355	1.238	0.232	0.207	0.181
E						
PLLE	0.944	0.886	0.830	0.083	0.058	0.037
KER1	7.979	0.936	1.202	0.424	0.082	0.042
KER2	0.440	0.359	0.313	0.041	0.023	0.014
F						
PLLE	1.021	0.957	0.859	0.088	0.058	0.038
KER1	1.631	0.811	0.553	0.141	0.054	0.026
KER2	0.635	0.585	0.506	0.095	0.068	0.045

Across all functional forms we note a clear decline in the PLLE based RMSEs corresponding to both $\hat{f}_0(q)$ and $\hat{f}_1(q)$ as T is allowed to increase. As expected from Proposition 1 however, the slope functions see their RMSEs decline substantially faster than their intercept counterparts. Within the same context it is also worth noting that the magnitude of the RMSEs is substantially larger for $\hat{f}_0(q)$ than for $\hat{f}_1(q)$. Another useful feature of the PLLE method is its robustness to alternative choices for k_{max} and different methods of obtaining an optimal k. We have experimented across a wide configuration of upperbounds and criteria (e.g., BIC, $\bar{R}^2$) and systematically obtained results that were both qualitatively and quantitatively similar to our estimates presented in Table 1.

It is now also interesting to compare our PLLE method with kernel based methods as proposed in Xiao (2009). A feature of the kernel based RMSEs presented in Table 1 is their sensitivity to the choice of a kernel function under small to moderate sample sizes. We note for instance that our approach often dominates the kernel based approach under the use of the quadratic kernel (denoted KER1) while the use of a uniform kernel (denoted KER2) leads to RMSEs that are sometimes smaller than the ones obtained under our PLLE approach. Under functions A, D and F our proposed PLLE approach compares favourably with both KER1 and KER2 for all sample size scenarios. Given also its trivial implementation via linear regression with dummy variables, our findings suggest that our proposed piecewise local linear approach is a useful and powerful addition to the toolkit on functional coefficient estimation, further supporting the conclusions in Banerjee (2007).

5 An application to stock prices and dividends

In this section we apply our Piecewise Local Linear estimation methodology to the study of linkages between stock prices and dividends. Since the early work of Campbell and Shiller (1987, 1988) it is well known that when prices and dividends are linked through a present value relationship it must be the case that they are cointegrated. This observation has generated a voluminous but often inconclusive empirical literature that aimed to document the presence or absence of linear cointegration between prices and dividends, often resulting in calls to extend the analysis to take into account potential nonlinearities as advocated by numerous theoretical models involving noise traders, bubble phenomena amongst numerous others [see Obstfeld and Rogoff (1986), Shleifer and Summers (1990) amongst many others]. It is indeed not uncommon to observe in the data episodes where the dynamics of prices appear disconnected from the underlying fundamentals making it difficult to view them as forming a linear equilibrium relationships when using detection methods designed for linear cointegration settings. One explanation that has often been put forward for explaining such occurences relies on the notion of a broadly defined sentiment concept [see Baker and Wurgler (2007)] with prices possibly disconnecting from fundamentals during episodes of particularly high sentiment. Our functional cointegration framework is particularly suitable for modelling such situations where the long run equilibrium relationship between two variables of interest may be affected by the dynamics of a third one.

Our empirical analysis uses monthly US data on real stock prices and dividends as provided on Robert Shiller's website (www.econ.yale.edu/shiller). Our functional cointegration setting aims to model a nonlinear cointegrating relationship between prices and dividends using as our functional variable the monthly consumer sentiment index compiled by the University of Michigan survey of consumers (data item UMCSENT downloaded from the St Louis Fed database). Since UMCSENT is available on a monthly basis from January 1978 onwards we restrict our analysis to the period 1978:1–2012:12. Due to the size of the available sample we implement our Piecewise Local Linear Estimation method allowing between k_{min}=20 and k_{max}=40 bins and selecting the optimal number via the AIC criterion. For comparison purposes we also implemented a Kernel based estimation approach using the uniform Kernel. Figure 2 below presents our estimated slope functional coefficient using the piecewise local linear and Kernel based methods. The latter has been implemented using a uniform kernel.

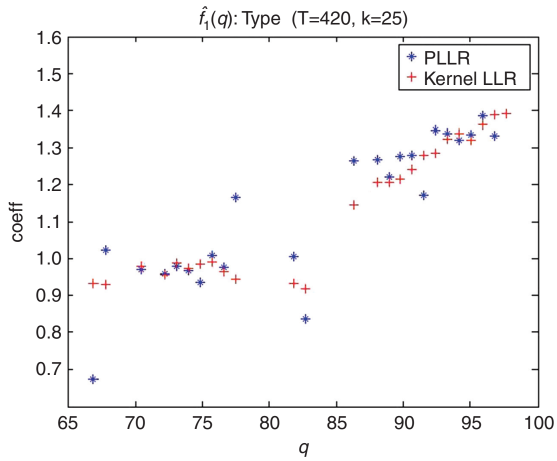


Figure 2 Sentiment based Functional Slope Coefficient Linking Prices and Dividends.

It is very interesting to note that for sentiment values below the low 80s (i.e., low or normal times as opposed to times of possible exuberance) the relationship between log stock prices and dividends appears to be linear and clustered along 1 as predicted by theory whereas for larger values of the sentiment index we observe a functional coefficient that is positively sloped and increasing in q. Note that our estimation method has been implemented using $q_t \equiv UMCSENT_t$ but results were virtually identical using its lagged value.

This interpretation, albeit informal is very much supportive of the idea that long run linkages between prices and fundamentals may be affected by sentiment. It is in fact possible to formalise our arguments by comparing the residuals of a linear versus functional fit. Implementing a standard Engle-Granger test on our monthly series of log stock prices and dividends we were unable to reject the null hypothesis of no cointegration. In a first instance the linear and potentially cointegrating regression linking p_t and d_t was given by $\hat{p}_t = 3.960 + 1.067\, d_t$. An ADF test on the associated residuals from this regression, say $\hat{u}_{L,t}$ subsequently led to a t-ratio of –1.888 with a corresponding p-value of 0.586 on the coefficient associated with $\hat{u}_{L,t-1}$. This clearly suggests that over the period 1978:1–2012:12 the data do not support the existence of a linear cointegrating relationship between real stock prices and dividends (in logs).

We next repeated the same exercise by considering the residuals from the functional regression, say $\hat{u}_{FC,t}$. The auxiliary ADF regression led to $\Delta\hat{u}_{FC,t} = -0.138\hat{u}_{FC,t-1} + lags$ with a t-ratio of –3.997 associated with the coefficient on $\hat{u}_{FC,t-1}$. Using the cointegration based critical values of the ADF statistic now leads to a rejection of the null at all conventional significance levels [see MacKinnon (2010)]. This is an interesting and powerful result. It suggests that log stock prices and dividends are linked via a linear or close to linear cointegrating relationship during normal times but the shape of the same relationship changes during certain episodes proxied by the level of a sentiment indicator.

At this stage it is also important to reiterate that our goal here was a simple illustration of our estimation methodolody using real data rather than conducting any formal follow up inferences such as testing for constancy of the coefficients or building confidence intervals for our functional coefficients. Doing so raises numerous technical challenges that would push us beyond the scope of this paper.

6 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. We subsequently used our methodology to illustrate the presence of an interesting relationship between stock prices and dividends driven by a sentiment indicator. 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.

Acknowledgement: We wish to thank seminar participants at the 2012 SNDE conference in Istanbul for very useful comments and suggestions. The second author also wishes to thank the ESRC for partial financial support.

APPENDIX

LEMMA 1: 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

$$\begin{aligned}&|E[(q_{t-1}-q_r)^m I_{rt-1}]|=|\int_{H_r}(q-q_r)^m g_q(q)dq|\\&\leq \int_{H_r}|q-q_r|^m g_q(q)dq\\&\leq h^m \int_{H_r} g_q(q)dq=const*h^{m+1}\end{aligned} \tag{8}$$

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

$$\begin{aligned}\hat{\beta}_{0r}&=\bar{y}_r-\hat{\beta}_{1r}\bar{x}_r\\ \hat{\beta}_{1r}&=\frac{\sum(x_t-\bar{x}_r)I_{rt-1}y_t}{\sum(x_t-\bar{x}_r)^2 I_{rt-1}}\end{aligned} \tag{9}$$

with $\bar{x}_r=\sum x_t I_{rt-1}/\sum I_{rt-1}$ and $\bar{y}_r=\sum y_t I_{rt-1}/\sum I_{rt-1}$. 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 $q\in H_r$

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

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

$$\begin{aligned}\hat{\beta}_{1r}-f_1(q)&=\frac{\sum(x_t-\bar{x}_r)I_{rt-1}[f_0(q_{t-1})+f_1(q_{t-1})x_t]}{\sum(x_t-\bar{x}_r)^2 I_{rt-1}}+\frac{\sum(x_t-\bar{x}_r)I_{rt-1}u_t}{\sum(x_t-\bar{x}_r)^2 I_{rt-1}}\\&=f_0'(q)\frac{\sum(x_t-\bar{x}_r)(q_{t-1}-q)I_{rt-1}}{\sum(x_t-\bar{x}_r)^2 I_{rt-1}}+f_1'(q)\frac{\sum x_t(x_t-\bar{x}_r)(q_{t-1}-q)I_{rt-1}}{\sum(x_t-\bar{x}_r)^2 I_{rt-1}}\\&+\frac{\sum(x_t-\bar{x}_r)I_{rt-1}u_t}{\sum(x_t-\bar{x}_r)^2 I_{rt-1}}.\end{aligned} \tag{10}$$

It is now also convenient to reformulate the above as

$$\begin{aligned} T\sqrt{h}(\hat{\beta}_{1r}-f_1(q)) &= f_0'(q)\left(\frac{\sum(x_t-\bar{x}_r)(q_{t-1}-q)I_{rt-1}/T^2h}{\sum(x_t-\bar{x}_r)^2 I_{rt-1}/T^2h}\right)T\sqrt{h}+ \\ & f_1'(q)\left(\frac{\sum x_t(x_t-\bar{x}_r)(q_{t-1}-q)I_{rt-1}/T^2h}{\sum(x_t-\bar{x}_r)^2 I_{rt-1}/T^2h}\right)T\sqrt{h}+ \\ & \frac{\sum(x_t-\bar{x}_r)I_{rt-1}u_t/Th}{\sum(x_t-\bar{x}_r)^2 I_{rt-1}/T^2h} \\ & \equiv T\sqrt{h}f_0'(q)A_r+T\sqrt{h}\,f_1'(q)B_r+C_r \end{aligned} \tag{11}$$

and the result follows by showing that $T\sqrt{h}\,A_r$ and $T\sqrt{h}\,B_r$ are asymptotically negligible when $Th^{3/2}\to 0$ while C_r is $O_p(1)$. Note that the denominators of the above are always bounded in distribution as $Th\to\infty$, since

$$\begin{aligned} & |\sum x_t^{\,2} I_{rt-1}/T^2h-g_q(q)\int B_v^2(s)| \\ \leq & |\sum x_t^{\,2} I_{rt-1}/T^2h-\sum B_v^2(t/T)I_{rt-1}/Th|+|\sum B_v^2(t/T)I_{rt-1}/Th-g_q(q)\int B_v^2(s)| \\ \leq & \sup_t |I_{rt-1}/h||\sum x_t^2/T^2-\sum B_v^2(t/T)/T|+(\sup_{s\in 0,1]} B_v(s)+1)^2|\sum I_{rt-1}/Th-g_q(q)|. \end{aligned} \tag{12}$$

Using the Markov inequality $\Pr(\sup_t |I_{rt-1}/h|>M)\leq \sup_t E(I_{rt-1})/Mh\leq \sup\, g_q(q)/M\to 0$ as $M\to\infty$ therefore I_{rt-1}/h is uniformly bounded. Our assumptions also ensure that $\sum x_t^{\,2}/T^2\Rightarrow\int_0^1 B_v^2$ [see Phillips (1987)] and finally the asymptotic negligibility of the last term in (12) as $Th\to\infty$ follows from a suitable law of large numbers for strong mixing processes [e.g., Hansen (1991), corollary 4]. See also [Hansen (2008, theorem 1)]. Similarly for $\bar{x}_r$.

We have for $q\in H_r, |q_{t-1}-q|<h$ and $f_1'(q)$ bounded,

$$\begin{aligned} T\sqrt{h}|B_r| &\leq T\sqrt{h}\frac{\sum|x_t(x_t-\bar{x}_r)(q_{t-1}-q)|I_{rt-1}}{\sum(x_t-\bar{x}_r)^2 I_{rt-1}} \\ &\leq Th^{3/2}\frac{\sum|x_t(x_t-\bar{x}_r)|I_{rt-1}}{\sum(x_t-\bar{x}_r)^2 I_{rt-1}}\to 0 \end{aligned} \tag{13}$$

since $Th^{3/2}\to 0$. The asymptotic negligibility of $T\sqrt{h}\,A_r$ follows along identical lines using the fact that

$$\begin{aligned} T\sqrt{h}|\sum(x_t-\bar{x}_r)(q_{t-1}-q)I_{rt-1}/T^2h| &\leq \sqrt{T}h^{3/2}\max_{t\leq T}\left|\frac{x_t}{\sqrt{T}}\right|\sum I_{rt-1}/Th \\ &\leq \sqrt{T}h^{3/2}\left(\sup_{s\in 0,1]} B_v(s)+1\right)\sum I_{rt-1}/Th \end{aligned} \tag{14}$$

since as before $(\sup_{s\in 0,1)} B_v(S)+1)\Sigma I_{rt-1}/Th$ is bounded $T\sqrt{h}\,A_r\to 0$.

Finally, for C_r, using $x_t=x_{t-1}+v_t$ we write

$$\frac{\sum(x_t-\bar{x}_r)I_{rt-1}u_t}{T\sqrt{h}}=\frac{\sum x_{t-1}I_{rt-1}u_t}{T\sqrt{h}}+\frac{\sum u_t v_t I_{rt-1}}{T\sqrt{h}}-\bar{x}_r\frac{\sum u_t I_{rt-1}}{\sqrt{Th}}. \tag{15}$$

Notice that $\Pr(|\sum u_t v_t I_{rt-1}/T\sqrt{h}|>\varepsilon)\leq\frac{1}{Th}E[u_t^2 v_t^2 I_{rt-1}]\to 0$. Same goes for the term $\sum u_t I_{rt-1}/\sqrt{Th}$ and $\bar{x}_r$ is bounded by $(\sup_{s\in 0,1)} B_v(S)+1)$ hence the third term is $O_p(1)$. So we can concentrate on $\sum x_{t-1}u_t I_{rt-1}/T\sqrt{h}$. We write as before

$$\left|\frac{1}{T\sqrt{h}}\sum x_{t-1}u_t I_{rt-1}\right|\leq\left(\sup_{s\in 0,1]} B_v(s)+1\right)\frac{1}{\sqrt{Th}}\sum|u_t|I_{rt-1}=O_p(1) \tag{16}$$

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

$$\hat{\beta}_{0r}-f_0(q)=f_0'(q)\frac{\sum(q_{t-1}-q)I_{rt-1}}{\sum I_{rt-1}}+f_1'(q)\frac{\sum(q_{t-1}-q)x_tI_{rt-1}}{\sum I_{rt-1}}+\frac{\sum u_tI_{rt-1}}{\sum I_{rt-1}}-\bar{x}_rO_p\left(\frac{1}{T\sqrt{h}}\right). \tag{17}$$

Applying suitable normalisations we reformulate (17) as

$$\sqrt{Th}(\hat{\beta}_{0r}-f_0(q))=f_0'(q)\left(\frac{\sum(q_{t-1}-q)I_{rt-1}}{\sum I_{rt-1}}\right)\sqrt{Th}+\left(f_1'(q)\frac{\sum(q_{t-1}-q)x_tI_{rt-1}}{\sum I_{rt-1}}\right)\sqrt{Th}+\frac{\sum u_tI_{rt-1}/\sqrt{Th}}{\sum I_{rt-1}/Th}+O_p(1). \tag{18}$$

Proceeding as above it is again straightforward to observe that the first two terms in the right hand side of (18) are asymptotically negligible while the third term is $O_p(1)$ by our Assumptions A.

References

Baker, M., and J. Wurgler. 2007. “Investor Sentiment in the Stock Market.” *Journal of Economic Perspectives* 21: 129–152.

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-Rico, 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-Rico, V., and J. Gonzalo. 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., Q. Li, and J. Y. Park. 2009. “Functional Coefficient Models for Nonstationary Time Series Data.” *Journal of Econometrics* 148: 101–113.

Campbell, J. Y., and R. J. Shiller. 1987. “Cointegration and Tests of Present Value Models.” *Journal of Political Economy* 95: 1602–1088.

Campbell, J. Y., and R. J. Shiller. 1988. “Stock Prices, Earnings, and Expected Dividends.” *Journal of Finance* 43: 661–676.

Chen, R., and R. S. Tsay. 1993. “Functional Coefficient Autoregressive Models.” *Journal of the American Statistical Association* 88: 298–308.

Cleveland, W. S., E. Grosse, and W. M. Shyu. 1991. “Local Regression Models.” In *Statistical Models in S* , edited by J. M. Chambers, and T.J. Hastie, 309–376. Pacific Grove, CA: Wadsworth & Brooks.

Fan, J., and W. Zhang. 1999. “Statistical Estimation in Varying Coefficient Models.” *The Annals of Statistics* 27: 1491–1518.

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

Gonzalo, J., and J. Pitarakis. 2006. “Threshold Effects in Cointegrating Relationships.” *Oxford Bulletin of Economics and Statistics* 68: 813–833.

Granger, C. W. J., and J. Hallman. 1991. “Nonlinear Transformations of Integrated Time Series.” *Journal of Time Series Analysis* 12: 207–234.

Hansen, B. 1991. “Strong Laws for Dependent Heterogeneous Processes.” *Econometric Theory* 7: 213–221.

Hansen, B. 2008. “Uniform Convergence Rates for Kernel Estimation with Dependent Data.” *Econometric Theory* 24: 726–748.

Hastie, T. J., and R. J. Tibshirani. 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., T. Myklebust, and D. Tjostheim. 2007. “Nonparametric Estimation in a Nonlinear Cointegration Type Model.” *Annals of Statistics* 35: 252–299.

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

MacKinnon, J. G. 2010. “Critical Values for Cointegrations Tests.” Queen’s Economics Department Working Paper No. 1227.

Obstfeld, M., and K. Rogoff. 1986. “Ruling Out Divergent Speculative Bubbles.” *Journal of Monetary Economics* 17: 349–362.
Phillips, P. C. B. 1987. “Time Series Regression with Unit Roots.” *Econometrica* 55: 277–302.
Saikkonen, P., and I. Choi. 2004. “Testing Linearity in Cointegrating Smooth Transition Regressions.” *Econometrics Journal* 7: 341–365.
Senturk, D., and H. G. Mueller. 2005. “Covariate-Adjusted Regression.” *Biometrika* 92: 75–89.
Senturk, D., and H. G. Mueller. 2006. “Inference for Covariate Adjusted Regression Via Varying Coefficient Models.” *Annals of Statistics* 34: 654–679.
Shleifer, A., and L. H. Summers. 1990. “The Noise Trader Approach to Finance.” *Journal of Economic Perspectives* 4: 19–33.
Wang, Q., and P. C. B. Phillips. 2009. “Structural Nonparametric Cointegrating Regression.” *Econometrica* 77: 1901–1948.
Xiao, Z. 2009. “Functional-Coefficient Cointegration Models.” *Journal of Econometrics* 152: 81–92.