Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

Some notes on nonlinear cointegration: A partial review with some novel perspectives

Dag Tjøstheim

To cite this article: Dag Tjøstheim (2020) Some notes on nonlinear cointegration: A partial review with some novel perspectives, Econometric Reviews, 39:7, 655-673, DOI: 10.1080/07474938.2020.1771900

To link to this article: https://doi.org/10.1080/07474938.2020.1771900

© 2020 The Author(s). Published with license by Taylor and Francis Group, LLC.

Published online: 17 Jun 2020.

Submit your article to this journal

Article views: 607

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=lecr20

ECONOMETRIC REVIEWS
2020, VOL. 39, NO. 7, 655–673
https://doi.org/10.1080/07474938.2020.1771900

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Some notes on nonlinear cointegration: A partial review with some novel perspectives

Dag Tjøstheim

Department of Mathematics, University of Bergen, Bergen, Norway

ABSTRACT

Some recent work on the analysis of nonlinear and nonstationary time series models is reviewed. A couple of novel results are obtained in extending nonlinear cointegrating regression models to a time series situation. All through the paper focus is on aspects that could lead to a more well-defined concept of nonlinear cointegration.

1. Introduction

Linear cointegration is now a well-established theory and a much-used technique in econometrics. The theory is founded by Clive Granger, see Granger (1983), for which he got the Nobel prize in economics. A basic paper is Engle and Granger (1987). Later developments are heavily influenced by Peter Phillips and by Søren Johansen; see e.g. Phillips (1991), Phillips and Solo (1992) and Johansen (1991).

Nonlinear cointegration is much less developed. It is not even entirely clear how it should be best defined. It is complicated because it involves nonlinearity as well as nonstationary. Further, it is multivariate and should retain an econometric interpretation analogous to linear cointegration. Peter Phillips and collaborators, in particular Park and Wang, have been main contributors to this subject using statistically advanced concepts like local time and nonlinear transformations of Brownian motion-like processes; see e.g. Park and Phillips (1999, 2001), Wang and Phillips (2009a,b), and Shi and Phillips (2012). A somewhat different approach using Markov theory of nonstationary processes has been advocated by Karlsen et al. (2007).

In the present paper I have roughly three objectives. First, I will try to sum up some recent developments, and try to explain why difficulties can and are encountered in extending linear concepts to nonlinear cointegration. In fact, in much of recent literature one has had to settle for a subclass of models, namely the so-called nonlinear cointegrating regression models. This field has been nicely surveyed in the book by Wang (2015). I will briefly review some current work here. My second objective is to extend a few results in Gao et al. (2015) and Li et al. (2016) from nonlinear cointegrating regression to the time series case, but I add that this may still not be sufficient to obtain a general nonlinear cointegration concept. A different and perhaps more satisfactory approach to this problem is taken in the special case of threshold cointegration, where I will briefly review work done by Cai et al. (2017). Third, I end up on a somewhat speculative note suggesting a novel local perspective in the nonlinear case. The details and indeed some main features remain to be worked out.

CONTACT Dag Tjøstheim Dag.Tjostheim@math.uib.no Department of Mathematics, University of Bergen, Bergen, Norway.

© 2020 The Author(s). Published with license by Taylor and Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

It should perhaps be stressed again at this point that the present paper is mainly a review paper. But the review is attempted made in such a way that it could possibly open up for a new and more appropriate notion of nonlinear cointegration. Throughout it is kept at a simple technical level with emphasis on ideas. However, it does contain a couple of new results, the technical details of which have been placed in an Appendix.

Here is a summary of the paper. Section 2 contains a very brief review of linear cointegration, stated with a view to extending this to the nonlinear case. Nonlinear error correction models and nonlinear cointegrating regression models, including functional-coefficient cointegration, are treated in Section 3. The nonstationary Markov chain approach is outlined in Section 4. Both the nonparametric and the parametric case are treated. In particular, in Section 4.2 results are extended from the parametric nonlinear cointegrating regression case to the time series situation. In Section 5 recent attempts by Cai et al. (2017) to define a threshold cointegration process, different from earlier approaches to do this, are looked at. In the last section a general concept of local nonlinear cointegration is sought introduced and compared to other local concepts such as local stationarity and local Gaussianity.

2. Vector autoregressive processes and linear cointegration

A collection of interdependent time series is often modeled by a vector AR (VAR) process

$$\mathbf{x}_t = \sum_{i=1}^{p} A_i \mathbf{x}_{t-i} + e_t \tag{1}$$

where $\mathbf{x}_t$ is m-dimensional and the matrices A_i are $m \times m$. The vector time series $\{\mathbf{x}_t\}$ is stationary if the roots of the characteristic polynomial $|A(z)| = |I_m - \sum_{i=1}^{p} A_i z^i|$ are outside the unit circle, that is, if $|A(z)| \neq 0$ for $|z| \leq 1$. Here I_m is the m-dimensional identity matrix, and $|\cdot|$ is the determinant. If there are k unit roots, say, and $m-k$ roots outside the unit circle, $\{\mathbf{x}_t\}$ is nonstationary. In the trivial and completely uninteresting case of independence between the component processes there are exactly k linear unit root I(1) processes and m – k stationary $I(0)$ processes. In the case of dependence between the component processes the k unit roots correspond to k common stochastic trends, and the m – k roots outside the unit circle lead to the existence of m – k linear combinations (eigenvectors corresponding to these roots) of the components which are stationary I(0) even though the component processes are nonstationary I(1). This property is called cointegration. The processes making up each of the m – k linear combinations move together in the long run. The cointegration concept was introduced in Granger (1983) and further developed in Engle and Granger (1987) and has spawned numerous papers.

There is a representation of a cointegrated system, the error correction representation, which has served as a basis for nonlinear extensions. The error correction representation is obtained by subtracting $\mathbf{x}_{t-1}$ from both sides of (1) and rearranging this equation as

$$\Delta \mathbf{x}_t = C\mathbf{x}_{t-1} + \sum_{i=1}^{p-1} B_i \Delta \mathbf{x}_{t-i} + e_t, \tag{2}$$

where Δ is the difference operator, where $C = -I_m + \sum_{i=1}^{p} A_i = -A(1)$ and $B_i = -\sum_{j=i+1}^{p} A_j, i = 1, \ldots, p-1$. When there are k unit roots of the characteristic polynomial, the matrix $C = -A(1)$ has rank $n = m - k$. The row space of C is then spanned by a basis of n linearly independent vectors, and we denote by β the $m \times n$ matrix whose columns form such a basis. Every row of C can now be written as a linear combination of the rows of β', the transpose of β. Thus, we can write $C = \alpha\beta'$, where α is an $m \times n$ matrix with full column rank, and Eq. (2) can then be written

$$\Delta \mathbf{x}_t = \alpha \mathbf{z}_{t-1} + \sum_{i=1}^{p-1} B_i \Delta \mathbf{x}_{t-i} + e_t \tag{3}$$

where $\mathbf{z}_{t-1} = \beta' \mathbf{x}_{t-1}$. One can solve for $\mathbf{z}_{t-1}$ obtaining

$$\mathbf{z}_{t-1} = (\alpha'\alpha)^{-1}\alpha' \left[\Delta \mathbf{x}_t - \sum_{i=1}^{p-1} B_i \Delta \mathbf{x}_{t-i} - e_t \right] \tag{4}$$

so that $\mathbf{z}_t$ is I(0). Thus, the linear combinations $\mathbf{z}_t = \beta' \mathbf{x}_t$ of nonstationary components are stationary, and the rows of β' are the cointegrating vectors. The term 'error correction' first appeared in Phillips (1957) and another pioneer was Sargan (1964). For an economic interpretation of the error correction representation, see Hamilton (1994, p. 581).

The error correction representation has been further developed in several papers; see e.g. Johansen (1988, 1991, 1995). The basis of these developments in the statistical literature is reduced rank regression.

The estimation and testing theory of linear cointegrating systems is now well developed. It makes systematic use of functional limit results and expansions in terms of the multidimensional Wiener process. Empirical aspects of linear cointegration are covered in Juselius (2006).

3. Attempts of a nonlinear generalization

3.1. Nonlinear error correction models

The error correction model is perhaps the one that has most often been used as a starting point for nonlinear extensions, usually with the nonlinear operation implemented only for the stationary process $\mathbf{z}_t$ in (3). We shall look at the nonlinear error correction model in this sub-section. In the next sub-section, we look at the more general problem of establishing nonlinear relationships directly on I(1) type variables.

We shall only consider the case of a bivariate process $\{\mathbf{x}_t\} = \{(x_{1t}, x_{2t})'\}$ in (2). If $\{x_{1t}\}$ and $\{x_{2t}\}$ are both I(1), then they are linearly cointegrated if there is a constant vector β such that $z_t = \beta' \mathbf{x}_t$ is I(0). It is generally true that if $\{z_t\}$ is stationary, then $\{g(z_t)\}$ is also stationary, assuming the mean and variance exist. A bivariate nonlinear error correction (NLEC) model extending (3) takes the form

$$\Delta \mathbf{x}_t = \boldsymbol{\alpha} g(z_{t-1}) + \sum_{i=1}^{p-1} B_i \Delta \mathbf{x}_{t-i} + \mathbf{e}_t$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ is a two-dimensional vector and g is a function such that $g(0) = 0$ and $Eg(z_t)$ exists. The function g can be estimated nonparametrically or by assuming a particular parametric form. Escribano (1986, 2004) used a cubic function of z_t in a UK money demand equation and achieved a parsimonious model.

An appealing form of NLEC models uses threshold error corrections. This device was originally introduced by Blake and Fomby (1997). The threshold error correction model has been further developed by several authors. Martens et al. (1998) treated the vector threshold error correction model TVEC, and Hansen and Seo (2002) provided a testing theory for the case where the cointegrating vector is estimated, and they treat a general multivariate case. More general switching mechanisms than the threshold one have been treated in Bec and Rahbek (2004). Saikkonen (2005) derived stability results for the general NLEC. Some additional references are Kapetanios et al. (2006), Kristensen and Rahbek (2010), Magri and Medeiros (2014), Seo (2011), Sjölander et al. (2017) and Wang et al. (2016a).

3.2. Parametric nonlinear regression with a nonstationary regressor ("nonlinear cointegrating regression")

It turns out to be considerably more tricky to obtain a general nonlinear cointegration relationship based on the basic formula (1). The starting point would then be given by a relationship of form

$$\mathbf{x}_t = \mathbf{g}(\mathbf{x}_{t-1}, \boldsymbol{\theta}) + \mathbf{u}_t \tag{5}$$

where $\mathbf{g}$ is a known vector function, and $\boldsymbol{\theta}$ is an unknown parameter vector and $\mathbf{u}_t$ is stationary. Let the series $\{\mathbf{x}_t\}$ be two-dimensional. Then this bivariate time series model is more difficult to treat in the nonstationary case than a regression relationship where $\{x_{1t}, x_{2t}\}$ is replaced by in general a nonstationary pair $\{y_t, x_t\}$ and

$$y_t = g(x_t, \boldsymbol{\theta}) + u_t \tag{6}$$

where the scalar function g is known, $\boldsymbol{\theta}$ is an unknown parameter vector and u_t is stationary. Of course (6) is of great general interest in time series regression analysis devoid of any cointegration interpretation.

For a stationary regressor x_t, the pair (x_t, y_t) in (6) can be analyzed using fairly standard methods. This is not the case in the nonstationary situation. The basic theory in that case was derived in two early much cited papers by Park and Phillips (1999, 2001). I give a short summary of their results in this section, because they are basic in this area, allows for introducing the concept of local time, and because they are possible to generalize.

Park and Phillips (1999, 2001) consider the rather general regression model (6), in which $\{u_t\}$ is a martingale increment process and $\{x_t\}$ an integrated process such that $\Delta x_t = v_t$. Here $\{v_t\}$ could be a moving average process or more generally a process such that

$$v_T(r) = \frac{1}{\sqrt{T}} \sum_{s=1}^{[rT]} v_s \tag{7}$$

converges to a Wiener process w_r. Here, $[rT]$ is the integer part of rT. Moreover, it is assumed that

$$(u_T(r), v_T(r)) \xrightarrow{d} (w_{1r}, w_{2r}) \tag{8}$$

where $\{(w_{1r}, w_{2r})'\}$ is a vector Wiener process, and

$$u_T(r) = \frac{1}{\sqrt{T}} \sum_{s=1}^{[rT]} u_s.$$

It should be noted that this set-up with $\{x_t\}$ being an I(1) type process excludes the possibility of analyzing the model

$$x_t = g(x_{t-1}, \boldsymbol{\theta}) + u_t \tag{9}$$

because the class of I(1) processes is not invariant under a general nonlinear transformation g. Then, since $\{x_t\}$ enters on both sides of the equality (9), it cannot in general be of I(1) type. This obstacle is a quite serious inconvenience in a cointegration set-up of the type we are striving for in the present paper. It will be taken up again in the next section using an alternative Markov process construction.

We will consider a least-squares estimator $\hat{\boldsymbol{\theta}}_T$ of $\boldsymbol{\theta}$ in (6); that is, $\hat{\boldsymbol{\theta}}_T$ is taken to minimize

$$Q_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \{y_t - g(x_t, \boldsymbol{\theta})\}^2 \tag{10}$$

Let $\dot{Q}_T = \partial Q_T/\partial \boldsymbol{\theta}$ and $\ddot{Q}_T = \partial Q_T/\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'$. The asymptotic analysis of $\hat{\boldsymbol{\theta}}_T$ takes as its starting point the Taylor expansion

$$\dot{Q}_T(\hat{\boldsymbol{\theta}}_T) = \dot{Q}_T(\boldsymbol{\theta}_0) + \ddot{Q}_T(\boldsymbol{\theta}_T)(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_T$ is an intermediate value determined by the mean value theorem. Using a scaling factor ν_T and the fact that $\dot{Q}(\hat{\boldsymbol{\theta}}_T) = 0$, this implies

$$\nu_T(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) = \left[\nu_T^{-1}\ddot{Q}_T(\boldsymbol{\theta}_0)\nu_T^{-1}\right]^{-1}\nu_T\dot{Q}_T(\boldsymbol{\theta}_0) + o_p(1).$$

It is seen that this leads to the evaluation of sums of type $\sum_t h_1(x_t, \boldsymbol{\theta}_0)$ and $\sum_t h_2(x_t, u_t, \boldsymbol{\theta}_0)$ for some functions h_1 and h_2 depending on the function g and its derivative. The evaluation of such sums is a crucial part of the analysis, and many of its aspects are covered in Park and Phillips (1999). In general, they show (Park and Phillips 1999) that for so-called regular functions (including continuous and piecewise continuous functions)

$$\frac{1}{T}\sum_t h\left(\frac{x_t}{\sqrt{T}}\right) \xrightarrow{d} \int_0^1 h(w_{2r})dr.$$

Park and Phillips (1999, 2001) consider altogether four classes of functions:

1. Integrable functions with the property that $\int_{-\infty}^{\infty} h(x)dx$ exists and is finite, so that $h(x) \to 0$ at a fast enough rate as $x \to \pm\infty$.
2. Asymptotic homogeneous functions, having the property
$$h(\lambda x) = k(\lambda)H(x) + R(x, \lambda) \tag{11}$$
where $H(x)$ is a homogeneous function so that $H(\lambda x) = k(\lambda)H(x)$ for some $k(\lambda), R(x, \lambda)$ is dominated by $H(x)$ when $|x|$ gets large.
3. Asymptotic exponential functions, where h grows to infinity with the speed of an exponential function.
4. Super exponential functions, where h grows to infinity faster than the simple exponential.

These classes of functions lead to rather different types of behavior for $\sum_t h(x_t)$, and unlike the linear and homogeneous case, integrals of functions of a Wiener process do not suffice. One needs to introduce the concept of local time of the Wiener process,

$$L(t, s) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon}\int_0^t I(|w_r - s| < \varepsilon)dr$$

where I is the indicator function. It may be noted that $L(t, s)$ is a random process in both t and s. It essentially measures the time that w_r spends close to s in the time interval $[0, t]$. It can be introduced in a much more general setting than the Wiener process. It can be made meaningful both for Markov processes and semi-martingales. Much of its importance stems from the so-called occupation time formula. It states that if h is locally integrable, then

$$\int_0^t h(w_r)dr = \int_{-\infty}^{\infty} h(s)L(t, s)ds. \tag{12}$$

which again is valid in a much more general setting, see for instance Revuz and Yor (1994).

It is interesting to note that the move from the analysis of stationary series to unit root processes required the introduction of mathematical techniques based on the Wiener process. Now moving further to the analysis of nonlinear transformations of unit root processes involves further new mathematics, involving local time.

Park and Phillips (1999) prove under some regularity conditions that if h is integrable and x_t is an integrated process such that $\Delta x_t = v_t$ with $\{v_t\}$ as in (7), then

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h(x_t) \xrightarrow{d} \left(\int_{-\infty}^{\infty} h(s)ds\right) L(1,0) \tag{13}$$

as $T \to \infty$.

This result means that $\sum_t h(x_t)$ spreads out at a rate of (is balanced by a scaling factor) $\sqrt{T}$. Moreover, the integrability of h implies that h tends to zero far out, and that asymptotically only observations at zero are exploited in the accompanying Wiener process as indicated by the local time variable $L(1, 0)$. On the other hand, the Wiener process spend most of its time outside the finite domain leading to a slow convergence rate. The behavior is very different in the homogeneous case. Indeed, again under some regularity conditions, if h is asymptotically homogeneous satisfying the decomposition (11), then

$$\frac{1}{Tk(\sqrt{T})}\sum_{t=1}^{T} h(x_t) \xrightarrow{d} \int_0^1 H(w_r)dr = \int_{-\infty}^{\infty} H(s)L(1,s)ds \tag{14}$$

as $T \to \infty$. The last equality follows from the occupation time formula (12).

For an integrable $g(x, \theta)$ with a scalar parameter θ in (6), results such as in (13), under regularity conditions stated in Park and Phillips (2001), lead to the following central limit theorem for $\hat{\theta}_T$ minimizing (10):

$$T^{1/4}(\hat{\theta}_T - \theta_0) \xrightarrow{d} (L(1,0)\int_{\infty}^{\infty} \dot{g}(s,\theta_0)^2 ds)^{-1/2} w_1$$

where $\dot{g} = \partial g/\partial\theta$, and where w_1, the Wiener process at time 1, is a standard normal random variable. The convergence rate is seen to be slower than the standard parametric convergence rate $T^{-1/2}$ of the stationary case. It comes from the scaling factor of $T^{1/2}$ in (13) and from a corresponding scaling factor $T^{1/4}$ for sums of type $\sum h(x_t)u_t$, with u_t defined in (6), for an integrable function h.

The analogous result for a homogeneous type $g(x, \boldsymbol{\theta})$ with a vector parameter $\boldsymbol{\theta}$ in (6) is given by (under a number of regularity conditions)

$$\sqrt{T}\dot{k}(\sqrt{T})'(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} (\int_0^1 \dot{H}(w_{2r}, \boldsymbol{\theta}_0)\dot{H}(w_{2r}, \boldsymbol{\theta}_0)'dr)^{-1} \int_0^1 \dot{H}(w_{2r}, \boldsymbol{\theta}_0)dw_{1r}. \tag{15}$$

Here H is the homogeneous part of $g(x, \boldsymbol{\theta})$ defined analogously to (11), $\dot{H} = \partial H/\partial\boldsymbol{\theta}$, k is the asymptotic order of $g(x, \boldsymbol{\theta})$ as in (11) (it may depend on $\boldsymbol{\theta}$) and $\dot{k}$ is defined as the corresponding asymptotic order of $\dot{g}(x, \theta)$, so that $\dot{g}(\lambda x, \boldsymbol{\theta}) = \dot{k}(\lambda)\dot{H}(x, \boldsymbol{\theta})$ asymptotically as x gets large. Finally, (w_{1r}, w_{2r}) is the pair of Wiener processes appearing in (8). In the scalar linear case $g(x, \theta) = \theta x, \dot{g}(x, \theta) = x$, such that $k(\lambda) = \dot{k}(\lambda) = \lambda$. This gives $\sqrt{T}\dot{k}(\sqrt{T}) = \sqrt{T}\sqrt{T} = T$ leading to the well-known unit root convergence rate of T^{-1}, faster than the standard stationary rate. It is also easy to check that the formula in (15) reduces to the standard formula in the linear case. There have been a number of follow-up papers, see e.g. Wang and Phillips (2009a), Wang and Phillips

(2009b), Wang and Chan (2014), Wang and Wang (2013). Many of these developments are surveyed in the recent book by Wang (2015).

But there are still several challenging problems here. As already mentioned, even though the regression relationship (6) is sometimes called a nonlinear cointegrating relationship, it does not really have the same symmetry in y and x as in the linear cointegrating case. For that purpose, transformations of both y_t and x_t have been examined. Granger and Hallman (1991) considered this problem. A more recent contribution is Goldstein and Stigum (2011) considering joint transformations of (y_t, x_t). Extending the analysis beyond the bivariate case presents another difficulty.

Saikkonen and Choi (2004) and Choi and Saikkonen (2004) consider estimation and testing of the model (6), where g is a smooth transition function. They apply another type of asymptotics, so-called triangular array asymptotics (cf. Andrews and McDermott 1995). In this kind of asymptotics, the actual sample size is fixed at T_0, say, and the model is embedded in a sequence of models depending on a sample size T which tends to infinity. The embedding is obtained by replacing the I(1) regressor x_t in (6)) by $(T_0/T)^{1/2}x_t$. This change leads to a central limit theorem for the least squares estimate $\hat{\theta}$ with rate $T^{-1/2}$ under some regularity conditions including a three times differentiability condition on the function g in (6). It is seen that the triangular array asymptotics is rather different from that used in Park and Phillips (1999, 2001). Inspecting the proofs it does not appear straightforward to extend them to a time series model such as (5). In Saikkonen and Choi (2004), there are both simulation experiments and applications to real data. The method appears to work well and may well deserve to be more widely investigated and applied than is the case presently.

3.3. Functional-coefficient regression

Functional-coefficient models have been used extensively in stationary time series analysis and regression; see e.g. Chen and Tsay (1993). Cai et al. (2009) and Xiao (2009) have extended this kind of modeling to the cointegrating regression framework and demonstrated how this can be used to introduce a cointegrating vector. Basically, their models take the form

$$y_t = \boldsymbol{\beta}(z_t)'\mathbf{x}_t + u_t$$

where $\{y_t\}$ is an observed scalar time series, $\{\mathbf{x}_t\}$ is a vector time series of observed covariates. Here, in general $\{\mathbf{x}_t\}$ is a nonstationary I(1) series, although Cai et al. (2009) also treats the case where $\{\mathbf{x}_t\}$ is stationary. The process $\{z_t\}$ is observed and can be stationary or nonstationary. It can be thought of as a covariate describing market and macroeconomics conditions, say. In both Cai et al. (2009) and Xiao (2009) z_t is taken to be a scalar. Finally, $\boldsymbol{\beta}$ is a smoothly varying vector function of z. In the special case that $\boldsymbol{\beta}(z_t) = \boldsymbol{\beta}$, the model reduces to a linear cointegration model with $\boldsymbol{\beta}$ playing the role of a cointegrating vector. The function $\boldsymbol{\beta}(\cdot)$ is estimated nonparametrically by kernel estimation or by local polynomial estimation. The asymptotic theory of this estimate is derived under various sets of precisely stated regularity conditions in both Cai et al. (2009) and Xiao (2009). A mixed normal is involved in the asymptotic distribution, but the convergence rate is highly dependent on the assumptions made. Under the assumption of a stationary $\{z_t\}$ and $\{\mathbf{x}_t\}$ the convergence rate is the traditional nonparametric rate of $(Th)^{-1/2}$ where h is the bandwidth used in the nonparametric estimation of $\boldsymbol{\beta}$. On the other hand if $\{z_t\}$ is I(1) and $\{\mathbf{x}_t\}$ stationary, Cai et al. (2009), not surprisingly, obtains the slower rate of $T^{-1/4}h^{-1/2}$. See also a similar situation in Section 4.1. Xiao (2009) considers the case where $\{z_t\}$ is stationary and $\{\mathbf{x}_t\}$ is I(1). The nonstationarity means that there is a more efficient spread of values of $\{\mathbf{x}_t\}$, and the result is super convergence of rate $T^{-1}h^{-1/2}$.

The approach to defining and treating a cointegrating vector is somewhat different for the two publications. Cai et al. (2009) states that if $\{y_t\}$ is I(0), and that part of $\{\mathbf{x}_t\}$ is I(1), then $\boldsymbol{\beta}(z_t)$ is thought of as a varying coefficient cointegrating vector, whereas Xiao (2009) takes $\boldsymbol{\beta}(z_t)$ as

cointegrating vector if $y_t - \boldsymbol{\beta}(z_t)'\mathbf{x}_t = u_t$ is stationary. Letting the cointegrating vector $\boldsymbol{\beta}$ be a function of z_t means that it is stochastically time varying. It may be argued, as in the introduction of Xiao (2009), that this may represent a more realistic approach to cointegration in modeling say asset prices and market fundamentals empirically. Another empirical example, where the functional-coefficient cointegration vector seems preferable, is highlighted in Wang et al. (2016b) dealing with quantity theory and inflation.

It is natural to test for stability of the cointegrating vector; i.e., test H_0: $\boldsymbol{\beta}(z) = \boldsymbol{\beta}$ against H_1: varying coefficient $\boldsymbol{\beta}$. Moreover, one can test, irrespective of whether $\boldsymbol{\beta}(z)$ is constant or not, for cointegration; i.e., to test whether H_0: $u_t = y_t - \boldsymbol{\beta}(z_t)'\mathbf{x}_t$ is stationary against the alternative H_1: y_t and $\mathbf{x}_t$ are not cointegrated such that $y_t - \boldsymbol{\beta}(z_t)'\mathbf{x}_t$ is I(1). Both tests are of importance in practice, and both tests are treated in Xiao (2009).

4. Markov modeling: an alternative approach

The Markov approach to modeling of nonstationary time series and regression models was presented in Karlsen and Tjøstheim (2001) and Karlsen et al. (2007). Since then there have been a number of follow-up papers. The approach has the advantage that both regression models analogous to (6) and time series models similar to (5) can be treated. But there is a price to be paid since the framework, at least at present, is more restrictive concerning the residual process u_t in (5), and essentially only nonlinear generalizations of I(1) type nonstationarity can be treated.

We start this section by discussing the nonlinear I(1) concept, then treat nonparametric estimation in Section 4.1 and finally some recent contributions to parametric estimation in Section 4.2.

The approach taken in Karlsen and Tjøstheim (2001) is that a generalized I(1) class containing both linear and nonlinear models is associated with the class of null recurrent Markov chains. The starting point is the simple random walk $x_t = x_{t-1} + e_t, t \geq 1,$ where $e_t \sim \text{iid}(\mu, \sigma^2)$ The two basic properties that Karlsen and Tjøstheim try to extend to a larger class of nonlinear I(1) type processes are (i) the persistence of the random walk (its nonstationarity); and (ii) the possibility of establishing central limit results, but not necessarily with convergence to the Wiener process.

The random walk is a linear process and a Markov chain. The Markov chain property also holds for the nonlinear generalization

$$x_t = g(x_{t-1}) + e_t \quad t \geq 1 \tag{16}$$

and such a process can be both stationary and nonstationary. If $|g(x)| \leq c|x|$ for some $c < 1$ when $|x|$ is large enough, then (Meyn and Tweedie 2009) there exists an initial distribution for x_0, so that $\{x_t\}$ becomes stationary if started with this distribution. On the other hand, if g is such that $\{x_t\}$ is explosive, e.g. $g(x) = x^2$, then $\{x_t\}$ is a transient Markov chain and g cannot be estimated, at least not nonparametrically. A crucial property for $\{x_t\}$ to have for it being possible to estimate g nonparametrically is that it should be recurrent. This means that if $x_s = x$ for a certain time point s, then the Markov chain $\{x_t\}$ is guaranteed to be in an arbitrary small neighborhood around x with probability one at a future time point; the process recurs or regenerates. We refer to Karlsen and Tjøstheim (2001) for more precise statements.

Under relatively weak regularity conditions, Karlsen and Tjøstheim derive a central limit theorem for sums of the type $\sum_{s=1}^{T} h(x_s)$ properly scaled, where h is a function satisfying some moment conditions. The key to this derivation is to use the recurrence property of the Markov chain to decompose the above sum as

$$\sum_{s=1}^{T} h(x_s) = \sum_{s=1}^{\tau_1} h(x_s) + \sum_{s=\tau_1+1}^{\tau_2} h(x_s) + \ldots + \sum_{s=\tau_n+1}^{T} h(x_s) \tag{17}$$

corresponding to the recurrence times $\tau_1, \tau_2, \ldots, \tau_n \leq T$; i.e., the time points of the regenerations of the chain. Clearly, $n = n(T) \to \infty$ as $T \to \infty$, but at a slower rate in the nonstationary case.

Due to the Markov property, the components $\sum_{s=\tau_{i-1}+1}^{\tau_i} h(x_s), i = 1, ..., n$ are independent and identically distributed, and this can be used to prove a central limit result under the additional assumption that the distribution of the recurrence time intervals $S_i = \tau_i - \tau_{i-1}$ should not have a too heavy tail. More specifically, $\Pr\{S_i > s\}$ is essentially of the order $s^{-\boldsymbol{\beta}}, 0 < \boldsymbol{\beta} < 1$, so that $ES_i^k < \infty$ for $k < \boldsymbol{\beta}$. This property is named β-null recurrence in Karlsen and Tjøstheim (2001). The random walk corresponds to $\boldsymbol{\beta} = 1/2$, as was established by Kallianpur and Robbins (1954).

The class of recurrent Markov chains is subdivided into positive and null recurrent chains, depending on whether the expected recurrence time ES_i is finite or not. The positive recurrent case has $ES_i < \infty$ ($\boldsymbol{\beta} = 1$ in the above) and corresponds to stationarity, whereas the null recurrent case can be associated with a nonlinear extension of I(1). A unit root AR(p) process can be cast as a p-dimensional Markov chain, and in Myklebust et al. (2012) it is shown that it is β-null recurrent with $\boldsymbol{\beta} = 1/2$ under weak assumptions. This paper also contains a characterization of vector autoregressive time series as to when they are β-null recurrent, recurrent but not β-null recurrent, and transient. But the null recurrent class is not restricted to linear processes, and it has the useful invariance property that if $\{x_t\}$ is null recurrent (β-null recurrent) then the transformed process $\{h(x_t)\}$ is null recurrent ($\boldsymbol{\beta}$-null recurrent) for an arbitrary one-to-one transformation h. As mentioned, such an invariance property does not hold for the "ordinary" I(1) class of processes.

4.1. Nonparametric estimation in a nonlinear cointegration type framework

Karlsen et al. (2007, 2010) consider nonparametric estimation in a nonlinear nonstationary environment which in some respects is wider than that of Park and Phillips (1999, 2001) but in other respects more narrow. The class of models is defined by

$$y_t = g(x_t) + u_t \tag{18}$$

where x_t is nonstationary and β-null recurrent as defined earlier in this section, u_t is a stationary infinite-order moving average process or a Markov chain. In contrast to the set-up in Section 3.2 one can now allow $y_t = x_{t+1}$, in which case $\{x_t\}$ is a Markov chain when $\{u_t\}$ is iid. Karlsen and Tjøstheim (2001) in fact discuss estimation in this case. In the following we phrase our discussion in terms of the model (18) because of its closeness to nonlinear cointegrating regression models just discussed. The function g is unknown and the task is to estimate it nonparametrically. Except for trivial choices of g (e.g., $g =$ constant), the process $\{y_t\}$ will be nonstationary, but it will not be β-null recurrent, as it is not even a Markov chain. The analysis in Karlsen et al. (2007) is carried out in two cases: the case in which $\{x_t\}$ and $\{u_t\}$ are independent, and the one in which dependence is allowed between them. At its present state the dependence modeling also requires a boundedness condition for $\{u_t\}$.

The function $g(x)$ in (18) is estimated nonparametrically using the Nadaraya-Watson estimator

$$\hat{g}(x) = \sum_{t=1}^{T} y_t K_h(x_t - x) / \sum_{t=1}^{T} K_h(x_t - x) \tag{19}$$

where $K_h(u) = h^{-1}K(h^{-1}u)$ is the kernel with bandwidth h. Karlsen et al. (2007) prove that

$$\left\{ h \sum_{t=1}^{T} K_h(x_t - x) \right\}^{1/2} \left\{ \hat{g}(x) - g(x) - \text{ bias term} \right\} \xrightarrow{d} \mathcal{N}(0, \sigma^2 \int K^2(s)ds) \tag{20}$$

as $T \to \infty$. Here, $\sigma^2 = \text{var}(u_t)$. The bias term tends to zero as $T \to \infty$, and it is explicitly given in Karlsen et al. (2007). The convergence of $\hat{g}(x)$ to $g(x)$ is slower than in the stationary case. This is easy to explain since the null recurrence of $\{x_t\}$ means that it takes more time for the process to return to a neighborhood of the point x, and it is the points in the neighborhood of x

which are used in the nonparametric estimation. Roughly speaking, the sample size is in effect reduced from T to T^{β} with $\boldsymbol{\beta} = 1/2$ if $\{x_t\}$ is a random walk. Then the rate of convergence for $\hat{g}(x)$ equals $T^{-1/4}h^{-1/2}$. For a fixed h this is seen to be the same rate as the parametric estimation rate of $\hat{\boldsymbol{\theta}}$ with an integrable function $g(x, \boldsymbol{\theta})$ in (6). The kernel function K plays the role of the integrable function in the nonparametric case. It should also be noted that in Karlsen et al. (2007) the so-called Mittag-Leffler process is the analogy of the local time process $L(t, 0)$. Wang and Phillips (2009a) and Wang and Phillips (2009b) use the local time as an alternative to obtain an asymptotic theory of the nonparametric estimates treated in Karlsen et al. (2007), but again this approach does not allow y_t to be replaced by x_{t+1} in (18). Finally, it should be noted that in contrast to the majority of limit theorems in that paper the limit in (20) is Gaussian. This is due to the stochastic scaling used.

A theory for specification testing is established in Gao et al. (2009a) for the time series regression case and in Gao et al. (2009b) for the time series autoregressive case. In the time series regression case a more general model and weakened assumptions are considered in Wang and Phillips (2012) using local time arguments.

4.2. Markov techniques in the parametric case

This sub-section is based on recent work by Li et al. (2016). It is of interest to extend the Markov framework of the nonparametric case to the parametric case. There are several reasons for that: (a) the parametric case has a much faster convergence rate and this should be exploited if one thinks that a parametric model will give a suitable approximation to the nonlinear structure, (b) it is of interest to compare the results that can be obtained in the parametric case using Markov arguments with corresponding results by Park and Phillips (1999, 2001) and others using their approach, (c) the Markov approach has a better chance of being extendable from the regression to the time series case, and I will in fact indicate how this can be done in the present paper. This last item is of some importance because it can be used as a starting point to a more "genuine" definition of nonlinear cointegration in the multivariate time series case.

I start by discussing the regression model and then go on to the time series model. A fundamental obstacle in trying to extend the Markov technique of Karlsen and Tjøstheim (2001) to the parametric case is their moment boundedness assumption A_0. This condition means that the two first moments of a typical term $\sum_{s=\tau_i+1}^{\tau_{i+1}} h(x_s)$ in (17) have to be bounded. For a parametric regression model this would be satisfied if the function g in $y_t = g(x_t, \boldsymbol{\theta})$ is bounded as a function of x, but this is not even fulfilled in the most trivial of cases, say, the linear model $y_t = ax_t + e_t$, where x_t is a simple random walk process. In the nonparametric approach of Section 4.1 this problem is not appearing since then the nonparametric estimate of g is given by (19) and the introduction of the kernel function K secures the needed boundedness of moments in the Markov decomposition (17).

In Li et al. (2016) this problem is solved by truncation, i.e. with a T-dependent threshold function M_T such that M_T tends to infinity with T. The model considered in that paper is

$$y_t = g(x_t, \boldsymbol{\theta}) + \sigma(x_t, \gamma)e_t$$

where heterogeneity is allowed. The main proof is carried out for the case of $\{e_t\}$ being iid, but it is extended under certain conditions to the case where $\{e_t\}$ is a linear process. The truncation is achieved through using a loss function

$$Q_{T,g}(\boldsymbol{\theta}) = \sum_{t=1}^{T} (y_t - g(x_t, \boldsymbol{\theta}))^2 1(|x_t \leq M_T) \tag{21}$$

and

$$Q_{T,g}(\gamma) = \sum_{t=1}^{T} \left[\log \bar{e}_t^2 - \log \sigma^2(x_t, \gamma)\right]^2 1(|x_t| \leq M_T)$$

where $\bar{e}_t$ are estimated residuals. The rate of M_T has in general a quite complicated form, but for a β-null recurrent Markov chain it can be taken as $M_T = M_0 T^{1-\beta} L_s(T)$ with the same notation as in Section 4.1 and where $L_s(\cdot)$ is a slowly varying function. As expected from Section 3.2, we get separate convergence rates for the integrable and the nonintegrable case. In the integrable case (no truncation needed) under some regularity conditions, the solution $\hat{\boldsymbol{\theta}}_T$ which minimizes the nontruncated loss function $Q_{T,g}(\boldsymbol{\theta})$ over the parameter space $\boldsymbol{\Theta}$ is consistent, i.e., $\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 = o_P(1)$, and has an asymptotically normal distribution of form

$$\sqrt{n(T)}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2 \ddot{L}^{-1}(\boldsymbol{\theta}_0)),$$

where $\sigma^2 = \mathrm{var}(e_t)$ and $\ddot{L}(\boldsymbol{\theta}_0) = \int \dot{g}(x, \boldsymbol{\theta}_0)\dot{g}^T(x, \boldsymbol{\theta}_0)\pi_s(dx)$ with π_s being the invariant measure of the null recurrent process $\{x_t\}$, and $n(T)$ is the number of regenerations in the interval $[0, T]$. Similarly, in the homogeneous case under another set of regularity conditions to be found in Li et al. (2016), we have consistency and

$$n^{1/2}(T)\mathbf{J}_g^{1/2}(T, \boldsymbol{\theta}_0)(\bar{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2 I_d)$$

where $\bar{\boldsymbol{\theta}}_T$ is the minimizer of (21), where $\mathbf{J}_g$ is a relatively complicated scaling factor, defined in Li et al. (2016), so that we get super convergence with the same rates as in the parametric case discussed in Section 3.2 when $\boldsymbol{\beta} = 1/2$. Moreover, I_d is the d-dimensional identity matrix, d being the dimension of $\boldsymbol{\theta}$. Corresponding results for the estimation of γ in the heterogeneous case are given in the paper. Park and Phillips (1999, 2001) also treat the case where $\{(u_t, v_t)\}$ in (6) and (8) are linear processes instead of Markov. In Li et al. (2016) it is shown how the just discussed results can in fact be generalized to cover the case where x_t is given by $x_t = x_{t-1} + u_t, u_t = \sum_{j=0}^{\infty} \phi_j e_{t-j}$, with $\{e_t\}$ being iid and $\{\phi_j\}$ satisfies some summability conditions and with $\phi = \sum_{j=1}^{\infty} \phi_j \neq 0$. Then $\{x_t\}$ is not Markov, but the Markov framework can nevertheless be used via the process $x_t^* = \phi \sum_{s=1}^{t} e_s$.

In Li et al. (2016) it is claimed that the framework can be used to extend the results to the time series case $x_t = g(x_{t-1}, \theta) + e_t$. I now endeavor to show this result. Actually, I also show as a by-product that the uniform consistency result for nonparametric regression estimation of Gao et al. (2015) can be extended to the time series case.

The proof in the regression case in Li et al. (2016) uses the set-up of the nonparametric Markov regression paper of Karlsen et al. (2007). In that paper independence is assumed between the regressor process $\{x_t\}$ and the innovation process $\{e_t\}$. This clearly does not hold in the time series case (5). On the other hand, in Karlsen and Tjøstheim (2001) that assumption is not used (and of course cannot be used), but rather one uses that e_t is independent of $\{x_s, s < t\}$. Following the argument in Karlsen and Tjøstheim (2001), in order to develop an asymptotic theory for the parameter estimation in the nonlinear autoregression (5), we need that the process $\{x_t\}$ is (Harris) null recurrent, but not that the compound process $\{x_t, e_{t+1}\}$ is also (Harris) null recurrent. This is because we essentially have to consider sums of products like $\dot{g}(x_t, \boldsymbol{\theta}_0)e_{t+1} = \dot{g}(x_t, \boldsymbol{\theta}_0)(x_{t+1} - g(x_t, \boldsymbol{\theta}_0))$, which are of the general form in Karlsen and Tjøstheim (2001) and can enter into a decomposition argument such as (17).

In extending the regression result in Li et al. (2016) to the time series case, in their assumption 3.1(ii) one has to replace the assumption that $\{x_t\}$ is independent of $\{e_t\}$ by the assumption that e_{t+1} is independent of $\{x_s, s \leq t\}$, which in fact follows from the way the time series (5) is

generated. Let us call this modified assumption 3.1(ii)′, and the corresponding modified 3.1 by 3.1′ in the following. No modifications of conditions 3.2 and 3.3 in Li et al. (2016) are needed. Then analogs of Theorems 3.1 and 3.2 of Li et al. (2016) with the obvious change that y_t is replaced by x_{t+1} can be stated. Since formally these are new results I write them out here:

Proposition 1 (*The integrable case*): *Let Assumption* 3.1′ *and* 3.2 *of* Li et al. (2016) *hold. Then*

(a) The solution $\hat{\boldsymbol{\theta}}_T$ which minimizes the loss function

$$Q_{T,g}(\boldsymbol{\theta}) = \sum_{t=1}^{T}(x_{t+1} - g(x_t, \boldsymbol{\theta}))^2$$

over the parameter space $\boldsymbol{\Theta}$ is consistent, i.e.,

$$\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta} = o_P(1).$$

(b) The estimator $\hat{\boldsymbol{\theta}}_T$ has the asymptotically normal distribution,

$$\sqrt{n(T)}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \sigma^2 \ddot{L}^{-1}(\boldsymbol{\theta}_0))$$

where $\sigma^2 = \mathrm{var}(e_t)$

Proposition 2 (The asymptotically homogeneous case): Let $\{x_t\}$ be a null recurrent Markov process such that Assumption 3.1(ii)′ and 3.3 of Li et al. (2016) hold. Then,

(a) The solution $\bar{\boldsymbol{\theta}}_T$ which minimizes the loss function

$$Q_{T,g}(\boldsymbol{\theta}) = \sum_{i=1}^{T}\left[x_{t+1} - g(x_t, \boldsymbol{\theta})\right]^2 I(|x_t| \le M_T)$$

over the parameter space $\boldsymbol{\Theta}$ is consistent, i.e.,

$$\bar{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 = o_P(1).$$

(b) The estimator $\bar{\boldsymbol{\theta}}_T$ has the asymptotically normal distribution

$$\sqrt{n(T)}\mathbf{J}_g^{1/2}(T, \boldsymbol{\theta})(\bar{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \sigma^2 \mathrm{I_d})$$

where $\mathbf{J}_g$ is defined in Li et al. (2016) and $\sigma^2 = \mathrm{var}(e_t)$.

The changes in the proofs relative to the proofs of Gao et al. (2015) and Li et al. (2016) are given in the Appendix. Note that we have only carried through the reasoning in the Appendix in the univariate $\{x_t\}$-case and with $\{e_t\}$ being iid. To discuss cointegration efficiently, similar analyses would have to be carried out in the multivariate, or at least the bivariate case. I believe that this is technically possible. But still one may possibly argue that this is only an exercise in asymptotic analysis for a multivariate nonstationary time series. It is not clear how a cointegrating vector and indeed cointegration should be introduced in this context. In an attempt to make headway on this problem we simplify the model and look at threshold cointegration models.

5. Threshold cointegration

A less ambitious goal is to restrict nonlinear cointegration to nonlinearity of threshold type. It is straightforward at least formally, to generalize the linear system

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{e}_t \tag{22}$$

to

$$\mathbf{x}_t = A\mathbf{x}_{t-1}1(\mathbf{x}_{t-1} \in D) + B\mathbf{x}_{t-1}1(\mathbf{x}_{t-1} \in D^c) + \mathbf{e}_t \tag{23}$$

where D is a subset of $\mathbb{R}^2$. Then linear cointegration concepts can be discussed separately for D and D^c, and one can have stationary like behavior in one region and cointegration like behavior in another, or even have different cointegration like behavior in each region. A major mathematical difficulty is that in general the $\{x_t\}$ process will be nonstationary, and at the same time x_t is the threshold variable determining the regions D and D^c. Earlier attempts have sought to avoid this problem by introducing a more indirect *stationary* threshold variable as mentioned in Section 3.1.

The above problem appears also in the univariate case if one tries to generalize the ordinary unit root process to a threshold unit root process. We refer to Gao et al. (2013) for a discussion and references.

Cai et al. (2017) start directly from the VAR representation (22) and then move to the threshold VAR representation (23) and make a first attempt of establishing a theory for such models. In doing so, they worked under some restrictions, many of which can probably be relaxed. First, they only look at the bivariate first order case. Second, the region D is supposed to be compact. In this region, quite general behavior of $\{x_t\}$ is permitted, that is, the matrix A in (23) can be arbitrary, including cases generating stationary and explosive type behaviors. In the region D^c, $\{x_t\}$ will be assumed to be a cointegrated process. Obviously, there are many generalizations in choosing regions; that is, there could be more than two regions, and in particular there could be more than one cointegrating region and with different cointegrating vectors.

Compared with the threshold error correction models in Section 3.1, the model given by (23) is self-excited in the traditional way such that a lagged variable constitutes the threshold variable.

A crucial tool in the analysis of (23) is again the theory of null recurrent Markov processes as developed in Karlsen and Tjøstheim (2001). The connection between linear VARs and null recurrent processes is pointed out in Myklebust et al. (2012). A first attempt to use this in a threshold context is given in Gao et al. (2013) in the univariate case. In that article, it is shown that the univariate threshold model considered there is β-null recurrent with $\boldsymbol{\beta} = 1/2$. That proof relies on a result in Meyn and Tweedie (2009), Proposition 11.5.4) giving conditions for univariate threshold processes to be recurrent. This result makes it possible to use the Markov splitting technique to obtain β-null recurrence. The proof in Meyn and Tweedie (2009) is based on a drift criterion using a Liapunov test function. Cai et al. (2017) essentially use the same technique in their proof of their main result, but a nontrivial change of test function is required. Once recurrence is established for (23), $\beta = 1/2$-null recurrence can be proved using a straightforward adaptation of the proof in Gao et al. (2013).

Under the assumption that B is a 2×2 matrix with eigenvalues 1 and c with $|c| < 1$, A is any 2×2 matrix, and some additional weak regularity conditions, the process $\{x_t\}$ defined by (23) is a 1/2-null recurrent process.

Similar to Lütkepohl (2005), Cai et al. (2017) first discuss the asymptotics for the transformed model obtained by diagonalizing the B matrix and then discuss the the asymptotics for the original model. This way of looking at the problem gives additional insight in the differences in convergence speed.

Using the matrix Q^{-1} composed of the eigenvectors of B, one can transform the process $\{\mathbf{x}_t\}$ into $\{\mathbf{y}_t = Q\mathbf{x}_t\}$ with,

$$\mathbf{y}_t = A_1\mathbf{y}_{t-1}1(\mathbf{y}_{t-1} \in C) + B_1\mathbf{y}_{t-1}1(\mathbf{y}_{t-1} \in C^c) + \varepsilon_t \tag{24}$$

where $B_1 = QBQ^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}, A_1 = QAQ^{-1}, \varepsilon_t = Q\mathbf{e}_t,$ and C is the compact set obtained by transforming D. The matrices A_1 and B_1 are estimated by ordinary least squares using observations in the respective regions C and C^c. As can be expected the convergence rate of $\hat{A}_1$ is slow and proportional to $n^{-1/2}(T),$ where, in turn, $n(T)$ is proportional to the number of visits to the compact set C. For the estimator of the B_1 matrix, similar to linear cointegration theory, the first column of the matrix estimator, corresponding to the unit root eigenvalue, converges with rate $T^{-1},$ whereas the second component corresponding to the eigenvalue c converges with rate $T^{-1/2}$. Transforming back to the A and B matrix, the estimate of the first one converges still with the slow rate $n(T)^{-1/2},$ whereas, due to the linear transformation from B_1 to B, all of the elements of $\hat{B}$ converges with rate $T^{-1/2}$. Details are given in Cai et al. (2017).

As in the ordinary linear cointegration, one can express the cointegrating vector in terms of the matrix $Q = (q_{ij})$. The Q^{-1} matrix consists of the eigenvectors $\mathbf{b}_1$ and $\mathbf{b}_2$ of B such that $B\mathbf{b}_1 = \mathbf{b}_1$ and $B\mathbf{b}_2 = c\mathbf{b}_2$ and $Q^{-1} = [\mathbf{b}_1\mathbf{b}_2]$. Then $QBQ^{-1} = B_1 = \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}$. The first component of $\{\mathbf{y}_t\}$ is a random walk in C^c, whereas the second component displays stationary behavior in C^c. Hence, since $y_t = Qx_t$, the cointegrating vector in this region is given by $[q_{21}q_{22}]$ such that $q_{21}x_{1,t} + q_{22}x_{2,t}$ behaves as a stationary process in C^c.

I refer to Cai et al. (2017) for the asymptotic theory of estimates of A and B and for the estimate of the cointegrating vector. The asymptotic distribution of the estimate of the cointegrating vector is nonstandard and is given in Theorem 6 of Cai et al. (2017), and the convergence rate is T^{-1}.

There are a number of extensions of the cointegration threshold model. For instance, in Cai et al. (2017) the threshold has been taken to be known in the general theory although estimation of the threshold is also considered in the unknown case and with good results. However, a good asymptotic theory for the threshold analogous to that developed in Chan (1993) is lacking. There is work in progress at this point.

6. Outlook

The threshold process is of course just a subclass of nonlinear time series. A natural question to ask is whether the nonlinear threshold cointegration can be extended to a wider class of nonlinear multivariate nonstationary processes. Is it possible to find a nonlinear extension where the cointegrating vector concept can still be made meaningful and perhaps be allowed to depend on a continuous index variable unlike the threshold case where there are a finite number of possible cointegrating vectors? Indeed, most of the models treated earlier in this paper such as nonlinear cointegrating regression processes and the time series version in (5) lack the concept of cointegrating vector.

The first problem, since we would like to keep the Markov chain framework, is to find a class of multivariate nonlinear possibly nonstationary processes where its member processes are still null recurrent or at least recurrent. An obvious smooth version of the threshold processes is the class of so-called STAR models (Teräsvirta et al. 2010, Chapter 3) which consists of smooth transition autoregressive models. These are Markov processes, but there is not a well-developed theory of recurrence for such processes. Even in the univariate case, for the special case of an exponential AR process

$$x_t = (a + be^{-\gamma x_{t-1}^2})x_{t-1} + e_t$$

with $\gamma > 0,$ to my knowledge, it is not known whether $\{x_t\}$ is null recurrent for $a = 1$. (The term $be^{-\gamma x^2}$ tends to zero as $|x| \to \infty,$ but the behavior of $\{x_t\}$ seems to depend on the sign of b). A

somewhat crude approach to the problem is to assume that $\{x_t\}$ behaves as a random walk outside a fixed threshold; i.e. $x_t = g(x_{t-1}) + e_t$ for $|x_{t-1}| < c$, and $x_t = x_{t-1} + e_t$ for $|x_{t-1}| \geq c$. Then null recurrence for $\{x_t\}$ can be proved as in the threshold case. This can obviously be extended to the multivariate case with $\mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{e}_t$ with the matrix A having one eigenvalue equal to 1 and the others less than one in absolute value, and where this representation is valid outside some compact region for $\mathbf{x}_{t-1}$, and where the behavior of $\{\mathbf{x}_t\}$ is more or less arbitrary inside the compact set; e.g. $\mathbf{x}_t = \mathbf{g}(\mathbf{x}_{t-1}) + \mathbf{e}_t$ for some function $\mathbf{g}$. Again, null recurrence can be proved as in Cai et al. (2017). A smooth transition model is obtained if $\mathbf{g}(\mathbf{x})$ approaches $A\mathbf{x}$ in some smooth fashion as $\mathbf{x}$ approaches the boundary of the compact set (but of course this is different from the traditional STAR models which are typically tending to a linear model as $\mathbf{x}$ tends to *infinity*.

The above at least demonstrates that it is possible to find continuous type models that fulfills the null recurrence property. In the following we will just assume that we have a model $\mathbf{x}_t = \mathbf{g}(\mathbf{x}_{t-1}) + \mathbf{e}_t$ which is recurrent, (null recurrent or positive recurrent). To handle the problem of defining cointegration and a cointegrating vector it is convenient to write this in a functional-coefficient form as in Chen and Tsay (1993), such that

$$\mathbf{x}_t = A(\mathbf{x}_{t-1})\mathbf{x}_{t-1} + \mathbf{e}_t \tag{25}$$

where $A(\cdot)$ is a possibly nonlinear matrix function, and where of course the VAR model is obtained as a special case of $A(\cdot)$ being a constant. This is different from the models of Section 3.3 which are functional-coefficient models of cointegrating regression type. The next step consists in a localizing this relationship. The relationship (25)) is replaced by a family of linear VAR processes $\{\mathbf{x}_t^{\mathbf{x}}\}$ indexed by the spatial coordinate $\mathbf{x}$ such that

$$\mathbf{x}_t^{\mathbf{x}} = A_{\mathbf{x}}\mathbf{x}_{t-1}^{\mathbf{x}} + \mathbf{e}_t^{\mathbf{x}}$$

where for each fixed $\mathbf{x}$, $A_{\mathbf{x}}$ is an m-dimensional matrix, and $\{\mathbf{e}_t^{\mathbf{x}}\}$ consists of iid variables. This construction is similar to the local stationarity construction of Dahlhaus (1997, 2001), where a locally stationary process can be thought of as being composed of a family of stationary processes, and to the construction of a local Gaussian approximation, Tjøstheim and Hufthammer (2013), Otneim and Tjøstheim (2017), where a multivariate density is thought of as being composed by a family of multivariate Gaussian distributions. See also the forthcoming book Tjøstheim et al. (2021) for further developments and references. The latter construction can be extended to VAR processes as being composed of a family of multivariate Gaussian AR processes. The threshold processes just treated can be considered as a special case of this where $A_{\mathbf{x}}$ is a piecewise step function of $\mathbf{x}$.

The key idea is now to realize that for a fixed $\mathbf{x}$, $\{\mathbf{x}_t^{\mathbf{x}}\}$ is an ordinary VAR process, and linear cointegration can be discussed and a cointegrating vector can be found by doing an eigen analysis in the ordinary way. Of course, for some $\mathbf{x}$-s, $\{\mathbf{x}_t^{\mathbf{x}}\}$ may well be stationary. The original process $\{\mathbf{x}_t\}$, in the special case that $\{\mathbf{e}_t^{\mathbf{x}}\}$ does not depend on $\mathbf{x}$, can be recovered as

$$\mathbf{x}_t = A_{\mathbf{x}_{t-1}}\mathbf{x}_{t-1} + \mathbf{e}_t.$$

As for the locally stationary case and the local Gaussian case, to do statistical inference in such a framework $A_{\mathbf{x}}$ cannot change too quickly, and one will have to introduce a kernel function and a corresponding bandwidth that can be allowed to tend to zero as the number of observations tend to infinity. Again the "effective" number of observations $n(T)$ can be expected to play a decisive role. Further, I believe that the approach of Johansen using the full information maximum likelihood analysis of a cointegrated system as outlined for example in Chapter 20 of Hamilton (1994) will be advantageous to use in an asymptotic inference of local estimates because that approach treat all of the components of $\{\mathbf{x}_t\}$ on the same basis, which would be important in a local analysis which is based on local correlation and not local regression. Finally, for a high dimensional

$\{\mathbf{x}_t\}$, the curse of dimensionality will be encountered in a local analysis. A possible way out might be to use a simplification analog to the one used in Otneim and Tjøstheim (2017), by requiring that in the matrix $A_{\mathbf{x}}$ the entry $a_{ij}(\mathbf{x})$ is only allowed to depend on the pairs of coordinates (x_i, x_j), so that $a_{ij}(\mathbf{x}) = a_{ij}(x_i, x_j)$.

Of course, the details of such an approach remain to be carried out, and they are not trivial, but to my mind this would be a more genuine nonlinear cointegration methodology (as compared for example to cointegrating regression). Further, it trivially reduces to the linear cointegration theory for the x-independent $A_{\mathbf{x}}$, with the restriction, though, that only I(1) type of nonstationarity is allowed due to the requirement of null recurrence in an asymptotic analysis.

Appendix

A.1. Proof of Propositions 1 and 2

Most of the steps in the proofs of Li et al. (2016) can be taken over directly and trivially. The derivation in that paper also uses material from Gao et al. (2015). What is needed is to amend those proofs in those two papers that use the independence of $\{x_t\}$ and $\{e_t\}$ and replace those arguments with arguments that only use that e_{t+1} is independent of $\{x_s, s \le t\}$.

Independence is explicitly used in the proof of Lemma B.3 of Gao et al. (2015), and in the (b)-part of Lemma A.1 of Li et al. (2016). Moreover, the Bernstein inequality is used several times in Gao et al. (2015), and in the Lemma A.2 in Li et al. (2016). In its traditional form the Bernstein inequality is stated for independent identically distributed variables, but as will be seen below, there are versions of the Bernstein result that are valid for dependent variables and that is within the dependence framework desired by us. Below it will be seen how these two results can be changed so that the procedures of Gao et al. (2015) and Li et al. (2016) can be extended to fit into the time series framework of the present paper.

A.2. Conditioning argument to avoid independence assumption of $\{x_t\}$ and $\{e_t\}$ in Lemma B.3 of Gao et al. (2015)

Following the notation of Lemma B.3 in Gao et al. (2015), with the adjustment for the time series case, we write: $z_{\tau_{k-1}+s} = |L_{h,x}(x_{\tau_{k-1}+s})e_{\tau_{k-1}+s+1}|$ for $1 \le s \le \tau_k - \tau_{k-1}$. Note that in the time series case the processes $\{x_t\}$ and $\{e_t\}$ are not independent, but we have $y_t = x_{t+1}$ where e_{t+1} is independent of $\{x_s, s \le t\}$. We have to evaluate

$$E\left[z_{\tau_{k-1}+1}^{l_1} \cdots z_{\tau_k}^{l_{\tau_k-\tau_{k-1}}}\right].$$

For $1 \le s \le \tau_k - \tau_{k-1}$, let $\mathcal{F}_{\tau_{k-1}+s}$ be the σ algebra generated by $\{x_{\tau_{k-1}+u}, 1 \le u \le s; \tau_{k-1}, \tau_k\}$. Note that $\mathcal{F}_{\tau_{k-1}+u} \subset \mathcal{F}_{\tau_{k-1}+v}$ for $1 \le u < v \le \tau_k - \tau_{k-1}$. Then, by taking successive conditional expectation, and exploiting the product structure,

$$E\left[z_{\tau_{k-1}+1}^{l_1} \cdots z_{\tau_k}^{l_{\tau_k-\tau_{k-1}}}\right] = E\left[E\left[\cdots E\left[z_{\tau_{k-1}+1}^{l_1} \cdots z_{\tau_k}^{l_{\tau_k-\tau_{k-1}}} | \mathcal{F}_{\tau_{k-1}+1}\right] \cdots | \mathcal{F}_{\tau_k}\right]\right].$$

Using the independence of $\{e_s, s > t\}$ of x_t, the inner conditional expectation is given by

$$E\left[z_{\tau_{k-1}+1}^{l_1} \cdots z_{\tau_k}^{l_{\tau_k-\tau_{k-1}}} | \mathcal{F}_{\tau_{k-1}+1}\right] = E|e_1|^{l_1} |L_{h,x}(x_{\tau_{k-1}+1}))|^{l_1} \cdot E\left[z_{\tau_{k-1}+2}^{l_2} \cdots z_{\tau_k}^{l_{\tau_k-\tau_{k-1}}} | \mathcal{F}_{\tau_{k-1}+1}\right].$$

We next condition

$$E|e_1|^{l_1} |L_{h,x}(z_{\tau_{k-1}+1})|^{l_1} E\left[z_{\tau_{k-1}+2}^{l_2} \cdots z_{\tau_k}^{l_{\tau_k-\tau_{k-1}}} | \mathcal{F}_{\tau_{k-1}+1}\right]$$

on $\mathcal{F}_{\tau_{k-1}+2}$ and obtain

$$E|e_1|^{l_1} E|e_1|^{l_2} |L_{h,x}(x_{\tau_{k-1}+1}))|^{l_1} |L_{h,x}(x_{\tau_{k-1}+2})|^{l_2} \cdot E\left[z_{\tau_{k-1}+3}^{l_3} \cdots z_{\tau_k}^{l_{\tau_k-\tau_{k-1}}} | \mathcal{F}_{\tau_{k-1}+2}\right]$$

We continue in this fashion until at the end we obtain the analogous expression to that stated toward the end of the proof of Lemma B.3.

A.3. Avoidance of independence in the Bernstein inequality

For a nonlinear AR(p) type process, we generally get p-dependence for the variables $z_k(s) = \sum_{\tau_i+1}^{\tau_{i+1}} h(x_s)$ entering into the split chain decomposition analogs to (17), or see Karlsen and Tjøstheim (2001). To simplify assume that we have an AR(1)type process; i.e., resulting in 1-dependence. The Bernstein inequality is used a couple of places in Gao et al. (2015) in the traditional way of iid variables and in Lemma A.2 in Li et al. (2016). But this is not a problem. Actually, the Bernstein inequality can be re-formulated by looking at its proof: Let $z = \frac{1}{n}\sum_{i=1}^{n} z_i$, where the variables $\{z_i\}$ have the same distribution, but are one-dependent. Then following the steps of the proof of the Bernstein inequality, it is not difficult to show that

$$P(|z| \geq \epsilon) \leq \exp\left\{-\frac{\epsilon^2}{2\mathrm{var}(z)\left[1+\frac{\epsilon}{\mathrm{var}(z)}\frac{C}{3}\right]}\right\}$$

where $|z_i| \leq C$. In our case, looking at the transition between line 2 and 3 in the proof of (B.44) in Gao et al. (2015),

$$z = \frac{1}{q}\sum_{k=1}^{q}(\bar{z}_k(s_j) - E[\bar{z}_k(s_j)]) \doteq \frac{1}{q}\sum_{k=1}^{q} z_k.$$

Because of 1-dependence and the Schwartz inequality,

$$\mathrm{var}(z) = \frac{1}{q^2}(q\cdot \mathrm{var}(z_k) + 2(q-1)\cdot \mathrm{cov}(z_k, z_{k+1}) \leq \frac{3q-2}{q^2}\mathrm{var}(z_k) \leq \frac{3q-2}{q^2}M_1/h,$$

where M_1 is a constant and h the bandwidth in the kernel function used in Gao et al. (2015). Inserting in the above modified Bernstein inequality and in line 3 of (B.44) of Gao et al. (2015), line 4 of that inequality becomes

$$\leq \sum_{j=1}^{Q_n}\sum_{q=C_1 n^{\beta}L_s(n)}^{C_2 n^{\beta}L_s(n)} 2\exp\left\{-\frac{q(\bar{\eta}_n/2)^2 h}{2\frac{3q-2}{q}M_1 + (2/3)h\kappa_n\bar{\eta}_n}\right\}$$

from which the rest of the proof can be done as before.

These results were obtained by changing results from Gao et al. (2015). The relevant results of Li et al. (2016) essentially rely on the original results of Gao et al. (2015). Referring to the amended results of Gao et al. (2015) the results in Li et al. (2016) can be changed so that they cover the time series case, more explicitly, the proof of the (b)-part of Lemma A1 of Li et al. (2016) can be changed using the conditioning argument just presented. Similarly, in the proof of part (b) of Lemma A2, again the conditioning argument can be used, and the amended Bernstein inequality can be used to handle the expression (D.4) in that proof.

Acknowledgments

I am grateful to the editor and two referees for a number of valuable and constructive comments on an earlier version.

References

Andrews, D. W. K., McDermott, C. J. (1995). Nonlinear econometric models with deterministically trending variables. *The Review of Economic Studies* 62(3):343–360. doi:10.2307/2298032

Blake, N. S., Fomby, T. B. (1997). Threshold cointegration. *International Economic Review* 38(3):627–645. doi:10.2307/2527284

Bec, F., Rahbek, A. (2004). Vector equilibrium correction models with nonlinear discontinuous adjustments. *The Econometrics Journal* 7(2):628–651. doi:10.1111/j.1368-423X.2004.00147.x

Cai, B., Gao, J., Tjøstheim, D. (2017). A new class of bivariate threshold cointegration models. *Journal of Business & Economic Statistics* 35(2):288–305. doi:10.1080/07350015.2015.1062385

Cai, Z., Li, Q., Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics* 148(2):101–113. doi:10.1016/j.jeconom.2008.10.003

Chan, K. S. (1993). Consistency and limiting distribution of a least squares estimator of a threshold autoregressive model. *The Annals of Statistics* 21(1):520–533. doi:10.1214/aos/1176349040

Chen, R., Tsay, R. S. (1993). Functional coefficient autoregressive models. *Journal of the American Statistical Association* 88(421):298–308. doi:10.2307/2290725

Choi, I., Saikkonen, P. (2004). Tests of linearity in cointegrating smooth transition regressions. *The Econometrics Journal* 7(2):341–365. doi:10.1111/j.1368-423X.2004.00134.x
Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* 25(1):1–37. doi:10.1214/aos/1034276620
Dahlhaus, R. (2001). A likelihood approximation for locally stationary processes. *The Annals of Statistics* 28(6):1762–1794. doi:10.1214/aos/1015957480
Engle, R. F., Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica* 55(2):251–276. doi:10.2307/1913236
Escribano, A. (1986). Identification and modeling of economic relationships in a growing economy. PhD thesis. , San Diego: University of California.
Escribano, A. (2004). Nonlinear error correction: The case of money demand in the United Kingdom (1878–2000). *Macroeconomic Dynamics* 8(1):76–116. doi:10.1017/S1365100504030019
Gao, J., King, M., Lu, Z., Tjøstheim, D. (2009a). Nonparametric specification testing for nonlinear time series with non stationarity. *Econometric Theory* 25(6):1869–1892. doi:10.1017/S0266466609990363
Gao, J., King, M., Lu, Z., Tjøstheim, D. (2009b). Specification testing in nonlinear and nonstationary time series autoregression. *The Annals of Statistics* 37(6B):3893–3928. doi:10.1214/09-AOS698
Gao, J., Kanaya, S., Li, D., Tjøstheim, D. (2015). Uniform consistency for nonparametric estimators in null recurrent processes. *Econometric Theory* 31(5):911–952. doi:10.1017/S0266466614000577
Gao, J., Tjøstheim, D., Yin, J. (2013). Estimation in threshold autoregressive models with a stationary and a unit root regime. *Journal of Econometrics* 172(1):1–12. doi:10.1016/j.jeconom.2011.12.006
Goldstein, H., Stigum, B. P. (2011). *Nonlinear Cointegration in Foreign Exchange Markets*. Manuscript, Oslo: Department of Economics, University of Oslo.
Granger, C. W. J. (1983). *Co-integrated Variables and Error Correcting Models*. Discussion Paper 83-13, San Diego: Department of Economics, University of California.
Granger, C. W. J., Hallman, J. J. (1991). Nonlinear transformations of integrated time series. *Journal of Time Series Analysis* 12(3):207–224. doi:10.1111/j.1467-9892.1991.tb00078.x
Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
Hansen, B. E., Seo, B. (2002). Testing for two-regime threshold cointegration in vector correction models. *Journal of Econometrics* 110(2):293–318. doi:10.1016/S0304-4076(02)00097-0
Johansen, S. (1988). Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control* 12(2–3):231–254. doi:10.1016/0165-1889(88)90041-3
Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59(6):1551–1580. doi:10.2307/2938278
Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
Juselius, K. (2006). *The Cointegrated VAR Model: Methodologies and Applications*. Oxford: Oxford University Press.
Kallianpur, G., Robbins, H. (1954). The sequence of sums of independent random variables. *Duke Mathematical Journal* 21(2):285–307. doi:10.1215/S0012-7094-54-02128-6
Kapetanios, G., Shin, Y., Snell, A. (2006). Testing for cointegration in nonlinear smooth transition error correction models. *Econometric Theory* 22(2):279–303. doi:10.1017/S0266466606060129
Karlsen, H., Myklebust, T., Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *The Annals of Statistics* 35(1):252–299. doi:10.1214/009053606000001181
Karlsen, H., Myklebust, T., Tjøstheim, D. (2010). Nonparametric regression estimation in a null recurrent time series. *Journal of Statistical Planning and Inference* 140(12):3619–3626. doi:10.1016/j.jspi.2010.04.029
Karlsen, H. A., Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series models. *The Annals of Statistics* 29(2):372–416. doi:10.1214/aos/1009210546
Kristensen, D., Rahbek, A. (2010). Likelihood-based inference in nonlinear error-correction models. *Journal of Econometrics* 158(1):78–94. doi:10.1016/j.jeconom.2010.03.010
Li, D., Tjøstheim, D., Gao, J. (2016). Estimation in nonlinear regression with Harris recurrent Markov chains. *The Annals of Statistics* 44(5):1957–1987. doi:10.1214/15-AOS1379
Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. New York: Springer.
Magri, R., Medeiros, M. C. (2014). Nonlinear error correction models with an application to commodity prices. *Brazilian Review of Econometrics* 33(2):145–170. doi:10.12660/bre.v33n22013.24116
Martens, M., Kofman, P., Vorst, T. (1998). A threshold error correction model for intraday futures and index returns. *Journal of Applied Econometrics* 13(3):245–263. doi:10.1002/(SICI)1099-1255(199805/06)13:3<245::AID-JAE480>3.0.CO;2-E
Meyn, S. P., Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. New York: Springer.
Myklebust, T., Karlsen, H. A., Tjøstheim, D. (2012). Null recurrent unit root processes. *Econometric Theory* 28(1):1–41. doi:10.1017/S0266466611000119

Otneim, H., Tjøstheim, D. (2017). The locally gaussian density estimator for multivariate data. *Statistics and Computing* 27(6):1595–1616. doi:10.1007/s11222-016-9706-6
Park, J. Y., Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 15(3):269–298. doi:10.1017/S0266466699153015
Park, J. Y., Phillips, P. C. B. (2001). Nonlinear regression with integrated time series. *Econometrica* 69(1):117–161. doi:10.1111/1468-0262.00180
Phillips, A. W. (1957). Stabilization policy and the forms of lagged responses. *The Economic Journal* 67(266): 265–277. doi:10.2307/2227788
Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica* 59(2):283–306. doi:10.2307/2938258
Phillips, P. C. B., Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics* 20(2):971–1001. doi:10.1214/aos/1176348666
Revuz, D., Yor, M. (1994). *Continuous Martingale and Brownian Motion*, 2nd ed. New York: Springer.
Saikkonen, P. (2005). Stability results for nonlinear error correction models. *Journal of Econometrics* 127(1):69–81. doi:10.1016/j.jeconom.2004.03.001
Saikkonen, P., Choi, I. (2004). Cointegrating smooth transition regressions. *Econometric Theory* 20(2):301–340. doi:10.1017/S0266466604202031
Sargan, J. D. (1964). Wages and prices in the United Kingdomml: A study in econometric methodology. In Hart, P. E., Mills, G., Whittaker, J. N. eds., *Econometric Analysis for National Economic Planning.* London: Butterworths.
Seo, M. H. (2011). Estimation in nonlinear error correction models. *Econometric Theory* 27:201–234.
Shi, X., Phillips, P. C. B. (2012). Nonlinear cointegrating regression under weak identification. *Econometric Theory* 28(3):509–547. doi:10.1017/S0266466611000648
Sjölander, P., Månsson, K., Shukur, G. (2017). A new nonlinear asymmetric cointegration approach using error correction models. *Communications in Statistics – Simulation and Computation* 46(2):1661–1668. doi:10.1080/03610918.2014.999087
Teräsvirta, T., Tjøstheim, D., Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*. Oxford: Oxford University Press.
Tjøstheim, D., Otneim, H., Støve, B. (2021). *Statistical modeling using Local Gaussian Approximation*. Elsevier, forthcoming.
Tjøstheim, D., Hufthammer, K. O. (2013). Local gaussian correlation: A new measure of dependence. *Journal of Econometrics* 172(1):33–48. doi:10.1016/j.jeconom.2012.08.001
Wang, M., Chan, N. H., Yau, C. Y. (2016a). Nonlinear error correction model and multiple-threshold cointegration. *Statistica Sinica* 26:1479–1498. doi:10.5705/ss.2014.219t
Wang, Q. (2015). *Limit Theorems for Nonlinear Cointegrating Regression*. London: World Scientific.
Wang, Q., Chan, N. (2014). Uniform convergence rates for a class of martingales with application in non-linear co-integrating regression. *Bernoulli* 20(1):207–230. doi:10.3150/12-BEJ482
Wang, Q., Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25(3):710–738. doi:10.1017/S0266466608090269
Wang, Q., Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica* 77:1901–1948.
Wang, Q., Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *The Annals of Statistics* 40(2):727–758. doi:10.1214/12-AOS975
Wang, Q., Wang, R. (2013). Nonparametric cointegrating regression with Nhh errors. *Econometric Theory* 29(1): 1–27. doi:10.1017/S0266466612000205
Wang, Y., Tu, Y., Chen, S. X. (2016b). Improving inflation prediction with quantity theory. *Economics Letters* 149: 112–115. doi:10.1016/j.econlet.2016.10.023
Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics* 152(2):81–92. doi:10.1016/j.jeconom.2009.01.008