MR4215856 62G08 91B84
Wang, Qiying [Wang, Qi Ying] (5-SYD-SM);
Phillips, Peter C. B. (1-YALE-NDM); **Kasparis, Ioannis** (CY-CYP-NDM)
Latent variable nonparametric cointegrating regression. (English summary)
Econometric Theory **37** (2021), *no. 1,* 138–168.

This article investigates statistical properties of a nonparametric cointegrating regression model with a latent regressor. Let (y_t, z_t) be two nonstationary variables that are linked according to the nonparametric regression model

$$y_t = g(z_t) + \eta_t, \quad t = 1, \dots, n,$$

where $g(\cdot)$ is an unknown function and η_t is a zero-mean stationary disturbance that may be correlated with z_t. Throughout it is assumed that the true regressor z_t is unobservable, whereas its proxy x_t is available. Such circumstances arise when z_t is measured with error or given in filtered forms as in the mixed-data-sampling (MIDAS) regression or Hodrick-Prescott (HP) filtering.

In this scenario, it is a common practice to use x_t in place of z_t to estimate $g(x)$ nonparametrically as

$$\widehat{g}(x) = \frac{\sum_{t=1}^n y_t K\{(x_t - x)h\}}{\sum_{t=1}^n K\{(x_t - x)/h\}},$$

where $K(\cdot)$ is a nonnegative kernel function and h is a bandwidth. To explore the limit theory of $\widehat{g}(x)$, observe that

$$\widehat{g}(x) = \frac{\sum_{t=1}^n \eta_t K\{(x_t - x)/h\}}{\sum_{t=1}^n K\{(x_t - x)/h\}} + \frac{\sum_{t=1}^n g(z_t) K\{(x_t - x)/h\}}{\sum_{t=1}^n K\{(x_t - x)/h\}}. \tag{1}$$

While the first term can be handled in a manner analogous to correctly specified cases, the second term involves $S_n := \sum_{t=1}^n g(z_t) K\{(x_t - x)/h\}$, the limit theory of which depends on the linkage between two nonstationary variables, (x_t, z_t), and n, and is new in the literature.

To obtain a general result on the limit behavior of S_n, let a_t be asymptotically cointegrated with x_t so that

$$a_t = \gamma_{nt} x_t + w_{nt} + u_t,$$

where γ_{nt} is a sample-size-dependent coefficient with a limit γ_0 (e.g., an estimated coefficient), w_{nt} is an asymptotically negligible misspecification error, and u_t is a stationary disturbance. Also let ξ_t be the innovation for x_t and $d_n := \sqrt{\mathrm{Var}(\sum_{t=1}^n \xi_t^2)}$. Then, it is demonstrated that for a class of continuous functions $f(\cdot)$, a given x and the bandwidth satisfying $h + d_n/(nh) \to 0$ as $n \to \infty$,

$$\frac{d_n}{nh} \sum_{t=1}^n f(a_t) K\left(\frac{x_t - x}{h}\right) \xrightarrow{d} E\{f(\gamma_0 x + u_1)\} L_Z(1, 0), \tag{2}$$

where $L_Z(t, x)$ is the continuous local time for a fractional Ornstein-Uhlenbeck process Z_t.

The limit behavior of $\widehat{g}(x)$ can be obtained by applying (2) to the second term of (1). Suppose that (x_t, z_t) has the linkage

$$x_t = \alpha_{nt} z_t + w_{nt} + u_t,$$

where $\max_{1\le k\le n}|\alpha_{nk}^{-1}-1|\to 0$ as $n\to\infty$, and x_t (and thus z_t) is correlated with η_t. Then, the first term of (1) is shown to be $O_p\{(nh^2)^{1/4}\}$, whereas reading f as g in (2) yields

$$\left\{\frac{d_n}{nh}\sum_{t=1}^{n}K\left(\frac{x_t-x}{h}\right),\frac{d_n}{nh}\sum_{t=1}^{n}g(z_t)K\left(\frac{x_t-x}{h}\right)\right\}\xrightarrow{d}\{L_Z(1,0),E\{g(x-u_1)\}L_Z(1,0)\}.$$

Therefore, applying the continuous mapping theorem to the second term of (1), we finally have

$$\widehat{g}(x)=O_p\{(nh^2)^{1/4}\}+E\{g(x-u_1)\}+o_p(1)\xrightarrow{p}E\{g(x-u_1)\}=:g_1(x).$$

The function $g_1(x)$ is a weighted average of g locally around x with weights determined by the stationary distribution of the measurement error u. Also observe that $\widehat{g}(x)$ is usually inconsistent because $g(x)\neq g_1(x)$ in general. However, when g is linear, $g(x)=g_1(x)$ holds and thus $\widehat{g}(x)$ becomes consistent. This corresponds to the case in which the true regressor z_t is measured with a stationary error in a linear cointegrating regression.

REVISED (December, 2021)

Current version of review. Go to earlier version. *Masayuki Hirukawa*