Economics Letters 162 (2018) 18–21

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Consistent estimator of nonparametric structural spurious regression model for high frequency data

Minsoo Jeong

Department of Economics, Yonsei University Wonju Campus, Wonju, Gangwon, 26493, Republic of Korea

HIGHLIGHTS

- For continuous-time regression models with nonstationary errors, we showed that conventional nonparametric estimators are not consistent.
- We proposed a new consistent nonparametric estimator.
- We derived the exact convergence rate of the conditional variance of our new estimator.

ARTICLE INFO

Article history:
Received 18 May 2017
Received in revised form 4 October 2017
Accepted 6 October 2017

JEL classification:
C14
C22
C51
G10

Keywords:
Nonparametric regression
Nonstationary error term
Structural spurious regression
Consistency
High frequency data

ABSTRACT

We propose a new nonparametric estimator for continuous-time regression models with nonstationary error terms. While other conventional nonparametric estimators such as the Nadaraya–Watson and local linear estimators are not consistent, our estimator achieves consistency and asymptotic normality.

1. Introduction

Unlike the cointegrating regression, nonstationary regression models with nonstationary errors have received limited attention in the literature. This is mainly because nonstationary errors mostly imply that the regression is spurious, and even if the model is non-spurious, the nonstationarity itself can often be avoided by simply differencing the data. Recent studies on various structural spurious regression models are well illustrated in, e.g., Choi et al. (2008), Trapani (2012) and Baltagi et al. (2017). Using differenced data, however, does not always solve the problem especially for nonlinear cases, since differencing the data changes the entire dependence structure. For example, if we have a regression model

$$Y_t = f(X_t) + U_t, \tag{1}$$

where f is a nonlinear function, then nonparametrically regressing $Y_t - Y_{t-1}$ on $X_t - X_{t-1}$ will give us some function of $X_t - X_{t-1}$, which is obviously different from our objective function f, a function of X_t. Therefore, when f is nonlinear, a simple first-differencing technique will not give correct information on the shape of f, which implies that we may need a new estimation method for such cases.

In this paper, we first show that conventional nonparametric estimators fail to work for the continuous-time nonstationary error regression model. So far, the analyses of these estimators were largely unavailable in the continuous-time framework, though they are already known to be inconsistent for a simple discrete-time $I(1)$ error case. Some recent papers, e.g., Benhenni and Rachdi (2006) and Zhang (2016), consider similar subjects, but their models are only remotely related to economic or financial analyses. However, utilizing the new limit theorems in Jeong and Park (2013), Aït-Sahalia and Park (2016) and Kim and Park (2017), we show that these conventional estimators are inconsistent in our continuous-time framework.

E-mail address: mssjong@yonsei.ac.kr.

https://doi.org/10.1016/j.econlet.2017.10.007

After showing the inconsistency of the conventional estimators, we then propose a new proper nonparametric estimator. We show that our estimator is indeed consistent, and obtain the convergence rate of its conditional error explicitly in terms of the model parameters. It turns out that the convergence rate depends on both the time span and sampling interval of the data, which implies that we can reduce the conditional variance by using frequently observed data, even when the total time span is not long enough. This could be considered a useful property of the estimator especially when analyzing nonstationary high frequency data, since estimators based on nonstationary data usually suffer from slow rates of convergence with respect to the total time span.

The rest of the paper is organized as follows. Section 2 presents our regression model and assumptions. Section 3 introduces various nonparametric estimators and provides their asymptotics as our main result. Concluding remarks follow in Section 4. Mathematical proofs are collected in the online supplement.

2. Model and assumptions

We consider the regression model in (1), where $f : \mathbb{R} \to \mathbb{R}$ is three times continuously differentiable, and X and U are martingale diffusions. This martingale condition is to simplify our analysis, but we expect that it can be extended to general diffusions without difficulties. Moreover, since any diffusion process becomes a martingale diffusion after scale transformation, as noted in Jeong and Park (2013), it is actually not a restriction of the model as long as we know the functional form of the scale function. For the specification of X and U, we let

$$dX_t = \sigma(X_t)dW_t, \qquad dU_t = \omega(X_t)dV_t$$

with $\mathbb{E}U_0 = 0$, where W and V are mutually independent standard Brownian motions. With this specification, the instantaneous volatility of U is allowed to be dependent on the current value of X, while the innovations of X and U are independent of each other. We let the domain of X be $\mathbb{R} = (-\infty, \infty)$ without loss of generality, since we only consider driftless null recurrent diffusions in this paper.

For the data, we suppose that X and Y are observed at intervals of length Δ over time $(0, T]$, which implies that the sample size is given by $n = T/\Delta$. We use the two dimensional asymptotics as illustrated in Aït-Sahalia and Park (2016), which allows that the sample span increases and the sampling interval decreases at the same time, i.e., $T \to \infty$ and $\Delta \to 0$.

Hereafter, we denote the kernel function as K and the bandwidth parameter as h. We also write $f \in RV_a$ when f is regularly varying with index a at the boundaries for some $a \in \mathbb{R}$. A function $\ell : (0, \infty) \to (0, \infty)$ is called slowly varying at ∞ if

$$\lim_{x\to\infty} \frac{\ell(\lambda x)}{\ell(x)} = 1$$

for all $\lambda > 0$, and $f : (0, \infty) \to (0, \infty)$ is called regularly varying with index a at ∞ if $f(x) = x^a \ell(x)$ for some slowly varying function ℓ. Since typical examples of slowly varying functions are the functions of logarithmic order or the functions with non-zero finite limits at the boundaries, a regularly varying function with index a can be roughly understood as a function whose boundary behavior is similar to the behavior of a power law function with exponent a. Therefore, for most of the regularly varying functions, the limit order of f can be written only by a power law function, though, for some cases in which the slowly varying component ℓ in $f(x) = x^a \ell(x)$ is the logarithmic function, for example, the logarithmic function still remains in the limit order of f. Readers are referred to Bingham et al. (1989) for the details of regularly varying functions, and to Kim and Park (2017) for an extension to functions defined on $\mathbb{R}$.

Now with these notations we assume the following.

Assumption 2.1. (a) $\sigma^2(x), \omega^2(x) > 0$ and are twice continuously differentiable for $x \in \mathbb{R}$.
(b) $\sigma^2 \in RV_{-r}$ for $r > -1$, and its limit order only consists of a power law function.
(c) $\omega^2 \in RV_{-s}$ for $s < r + 1$, and its limit order only consists of a power law function.
(d) $K : \mathbb{R} \to [0, \infty)$ is bounded, twice continuously differentiable, and symmetric with respect to the origin. Also, K has a bounded support and satisfies $\int_{\mathbb{R}} K(u)du = 1$.
(e) $\Delta T^4 \to 0$ and $h \to 0$ as $T \to \infty$ and $\Delta \to 0$.

The conditions in Assumption 2.1(a) are standard regularity conditions for the diffusion processes. Assumption 2.1(b) implies that σ^2 is asymptotically homogeneous of degree $-r$ at the boundaries of $\mathbb{R}$, where the condition $r > -1$ requires that X is not stationary, as noted in Jeong and Park (2013). Since we only consider nonstationary processes here, $r > -1$ is intentionally imposed and does not restrict the model. The absence of nontrivial slowly varying components in the limit orders of σ^2 and ω^2 is not absolutely necessary, but we impose this to write the convergence rates of the estimators only as functions of T and r. This will help clarify the exposition of our result.

In Assumption 2.1(c), it seems that the condition $s < r + 1$ can be relaxed, but we impose this to utilize the limit theorems in Jeong and Park (2013). Though it certainly restricts the model, our analysis still includes most of the common interesting cases such as $r = s$, in which both the regressor and the error term share the same asymptotic order. Therefore, we expect that it does little harm delivering the message of this paper. Assumption 2.1(d) is a standard condition on the kernel function. Assumption 2.1(e) is a condition to control the remainder terms arising in our asymptotic expansions. This requires that the sampling interval diminishes fast enough as the total data span increases, which implies that our asymptotics is more suitable for analyzing high frequency data.

3. Estimators and their asymptotics

In this section, we will show that the Nadaraya–Watson and local linear estimators are not consistent, while our new estimator achieves consistency. Hereafter we let $K_h(z) = K(z/h)/h$ for notational simplicity.

3.1. Nadaraya–Watson estimator

The Nadaraya–Watson estimator of f at each $x \in \mathbb{R}$ is given by

$$\hat{f}_{NW}(x) = \frac{\sum_{i=1}^n K_h(X_{i\Delta} - x)Y_{i\Delta}}{\sum_{i=1}^n K_h(X_{i\Delta} - x)}.$$

To analyze the asymptotics of the estimator, we decompose $\hat{f}_{NW}(x)$ such that $\hat{f}_{NW}(x) = \hat{f}^m_{NW}(x) + \hat{f}^e_{NW}(x)$, where

$$\hat{f}^m_{NW}(x) = \frac{\sum_{i=1}^n K_h(X_{i\Delta} - x)f(X_{i\Delta})}{\sum_{i=1}^n K_h(X_{i\Delta} - x)},$$

$$\hat{f}^e_{NW}(x) = \frac{\sum_{i=1}^n K_h(X_{i\Delta} - x)U_{i\Delta}}{\sum_{i=1}^n K_h(X_{i\Delta} - x)}.$$

Here, $\hat{f}^m_{NW}(x)$ represents the conditional mean part of the estimator, while $\hat{f}^e_{NW}(x)$ represents the residual net of the conditional mean. In the following theorem, we provide asymptotics for each term separately.

Theorem 3.1. *Under Assumption 2.1, we have*

$$\hat{f}^m_{NW}(x) \to_p f(x), \qquad \frac{1}{\sqrt{T^{(r-s+2)/(r+2)}}}\hat{f}^e_{NW}(x) \to_d \mathbb{MN}\big(0, \Omega_{NW}\big)$$