Journal of Multivariate Analysis 153 (2017) 121–135

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Nonparametric estimation of the distribution of the autoregressive coefficient from panel random-coefficient AR(1) data

Remigijus Leipus [a,b], Anne Philippe [c,d], Vytautė Pilipauskaitė [b,c,*], Donatas Surgailis [b]

[a] Vilnius University, Faculty of Mathematics and Informatics, Naugarduko 24, LT-03225 Vilnius, Lithuania
[b] Vilnius University, Institute of Mathematics and Informatics, Akademijos 4, LT-08663 Vilnius, Lithuania
[c] Université de Nantes, Laboratoire de Mathématiques Jean Leray, 44322 Nantes Cedex 3, France
[d] ANJA INRIA Rennes Bretagne Atlantique, France

A R T I C L E   I N F O

Article history:
Received 10 July 2015
Available online 23 September 2016

AMS 2010 subject classifications:
62G10
62M10
62G07

Keywords:
Random-coefficient autoregression
Empirical process
Kolmogorov–Smirnov statistic
Goodness-of-fit testing
Kernel density estimator
Panel data

A B S T R A C T

We discuss nonparametric estimation of the distribution function $G(x)$ of the autoregressive coefficient $a \in (-1, 1)$ from a panel of N random-coefficient AR(1) data, each of length n, by the empirical distribution function of lag 1 sample autocorrelations of individual AR(1) processes. Consistency and asymptotic normality of the empirical distribution function and a class of kernel density estimators is established under some regularity conditions on $G(x)$ as N and n increase to infinity. The Kolmogorov–Smirnov goodness-of-fit test for simple and composite hypotheses of Beta distributed a is discussed. A simulation study for goodness-of-fit testing compares the finite-sample performance of our nonparametric estimator to the performance of its parametric analogue discussed in Beran et al. (2010).

1. Introduction

Panel data can describe a large population of heterogeneous units/agents which evolve over time, e.g., households, firms, industries, countries, stock market indices. In this paper we consider a panel where each individual unit evolves over time according to order-one random coefficient autoregressive model (RCAR(1)). It is well known that aggregation of specific RCAR(1) models can explain long memory phenomenon, which is often empirically observed in economic time series (see [9] for instance). More precisely, consider a panel $\{X_i(t),\ t = 1, \ldots, n,\ i = 1, \ldots, N\}$, where each $X_i = \{X_i(t), t \in \mathbb{Z}\}$ is an RCAR(1) process with $(0, \sigma^2)$ noise and random coefficient $a_i \in (-1, 1)$, whose autocovariance

$$\mathrm{E}X_i(0)X_i(t) = \sigma^2 \int_{-1}^{1} \frac{x^{|t|}}{1-x^2}\, \mathrm{d}G(x) \tag{1.1}$$

* Corresponding author at: Vilnius University, Institute of Mathematics and Informatics, Akademijos 4, LT-08663 Vilnius, Lithuania.
E-mail address: vytaute.pilipauskaite@gmail.com (V. Pilipauskaitė).

http://dx.doi.org/10.1016/j.jmva.2016.09.007
0047-259X/

is determined by the distribution function $G(x) = \Pr(a \le x)$ of the autoregressive coefficient. Granger [9] showed, for a specific Beta-type distribution $G(x)$, that the *contemporaneous* aggregation of independent processes $\{X_i(t)\}$, $i = 1, \ldots, N$, results in a stationary Gaussian long memory process $\{\mathcal{X}(t)\}$, i.e.,

$$N^{-1/2} \sum_{i=1}^{N} X_i(t) \to_{\rm fdd} \mathcal{X}(t) \quad \text{as } N \to \infty, \tag{1.2}$$

where the autocovariance $\mathrm{E}\mathcal{X}(0)\mathcal{X}(t) = \mathrm{E}X_1(0)X_1(t)$ decays slowly as $t \to \infty$ so that $\sum_{t\in\mathbb{Z}} |\mathrm{E}\mathcal{X}(0)\mathcal{X}(t)| = \infty$.

A natural statistical problem is recovering the distribution $G(x)$ (the frequency of a across the population of individual AR(1) 'microagents') from the aggregated sample $\{\mathcal{X}(t),\ t = 1, \ldots, n\}$. This problem was treated in [5,6,12]. Some related results were obtained in [4,10,11]. Albeit nonparametric, the estimators in [5,12] involve an expansion of the density $g = G'$ in an orthogonal polynomial basis and are sensitive to the choice of the tuning parameter (the number of polynomials), being limited in practice to very smooth densities g. The last difficulty in estimation of G from aggregated data is not surprising due to the fact that aggregation *per se* inflicts a considerable loss of information about the evolution of individual 'micro-agents'.

Clearly, if the available data comprises evolutions $\{X_i(t),\ t = 1, \ldots, n\}$, $i = 1, \ldots, N$, of all N individual 'micro-agents' (the panel data), we may expect a much more accurate estimate of G. Robinson [15] constructed an estimator for the moments of G using sample autocovariances of X_i and derived its asymptotic properties as $N \to \infty$, whereas the length n of each sample remains fixed. Beran et al. [1] discussed estimation of two-parameter Beta densities g from panel AR(1) data using maximum likelihood estimators with unobservable a_i replaced by sample lag 1 autocorrelation coefficient of $X_i(1), \ldots, X_i(n)$ (see Section 6), and derived the asymptotic normality together with some other properties of the estimators as N and n tend to infinity.

The present paper studies nonparametric estimation of G from panel random-coefficient AR(1) data using the empirical distribution function:

$$\widehat{G}_{N,n}(x) := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\widehat{a}_{i,n} \le x), \quad x \in \mathbb{R}, \tag{1.3}$$

where $\widehat{a}_{i,n}$ is the lag 1 sample autocorrelation coefficient of X_i, $i = 1, \ldots, N$ (see (3.3)). We also discuss kernel estimation of the density $g(x) = G'(x)$ based on smoothed version of (1.3). We assume that individual AR(1) processes X_i are driven by identically distributed shocks containing both common and idiosyncratic (independent) components. Consistency and asymptotic normality as $N, n \to \infty$ of the above estimators are derived under some regularity conditions on $G(x)$. Our results can be applied to test goodness-of-fit of the distribution $G(x)$ to a given hypothesized distribution (e.g., a Beta distribution) using the Kolmogorov–Smirnov statistic, and to construct confidence intervals for $G(x)$ or $g(x)$.

The paper is organized as follows. Section 2 obtains the rate of convergence of the sample autocorrelation coefficient $\widehat{a}_{i,n}$ to a_i, in probability, the result of independent interest. Section 3 discusses the weak convergence of the empirical process in (1.3) to a generalized Brownian bridge. The Kolmogorov–Smirnov goodness-of-fit test for simple and composite hypotheses of Beta distributed a is discussed in Section 4. In Section 5 we study kernel density estimators of $g(x)$. We show that these estimates are asymptotically normally distributed and their mean integrated square error tends to zero. A simulation study of Section 6 compares the empirical performance of (1.3) and the parametric estimator of [1] to the goodness-of-fit testing for $G(x)$ under null Beta distribution. The proofs of auxiliary statements can be found in the Appendix.

In what follows, C stands for a positive constant whose precise value is unimportant and which may change from line to line. We write $\to_{\rm p}$, $\to_{\rm d}$, $\to_{\rm fdd}$ for the convergence in probability and the convergence of (finite-dimensional) distributions respectively, whereas $\Rightarrow$ denotes the weak convergence in the space $D[-1, 1]$ with the supremum metric.

2. Estimation of random autoregressive coefficient

Consider an RCAR(1) process

$$X(t) = aX(t-1) + \zeta(t), \quad t \in \mathbb{Z}, \tag{2.1}$$

where innovations $\{\zeta(t)\}$ admit the following decomposition:

$$\zeta(t) = b\eta(t) + c\xi(t), \quad t \in \mathbb{Z}, \tag{2.2}$$

where random sequences $\{\eta(t)\}$, $\{\xi(t)\}$ and random coefficients a, b, c satisfy the following conditions:

Assumption A_1. $\{\eta(t)\}$ are independent identically distributed (i.i.d.) random variables (r.v.s) with $\mathrm{E}\eta(0) = 0$, $\mathrm{E}\eta^2(0) = 1$, $\mathrm{E}|\eta(0)|^{2p} < \infty$ for some $p > 1$.

Assumption A_2. $\{\xi(t)\}$ are i.i.d. r.v.s with $\mathrm{E}\xi(0) = 0$, $\mathrm{E}\xi^2(0) = 1$, $\mathrm{E}|\xi(0)|^{2p} < \infty$ for the same p as in A_1.

Assumption A_3. b and c are possibly dependent r.v.s such that $\Pr(b^2 + c^2 > 0) = 1$ and $\mathrm{E}b^2 < \infty$, $\mathrm{E}c^2 < \infty$.

Assumption A_4. $a \in (-1, 1)$ is a r.v. with a distribution function (d.f.) $G(x) := \Pr(a \le x)$ supported on $[-1, 1]$ and satisfying

$$\mathrm{E}\Big(\frac{1}{1-|a|}\Big) = \int_{-1}^{1} \frac{\mathrm{d}G(x)}{1-|x|} < \infty. \tag{2.3}$$