AMSE JOURNALS-AMSE IIETA publication-2017-Series: Advances A; Vol. 54; N°2; pp 322-332
Submitted Mar. 31, 2017; Revised Jun. 11, 2017, Accepted Jun. 18, 2017
https://doi.org/10.18280/ama_a.540214

A Result of Convergence about Weighted Sum for Exchangeable Random Variable Sequence in the Errors-in-Variables Model

Mei Yang

Chongqing College of Electronic Engineering, Chongqing 401331, China
(meizi11106@163.com)

Abstract

This thesis discusses linear EV (errors-in-variables) regression models, that is, regression models with measurement errors. Because in practice, data are often obtained with measurement errors, EV model is more fit for application than the ordinary regression model. However, it is more complicated in the statistical inference and analysis, so research about this theory is very difficult. Due to the application of statistics, when the weight function uses real variables in EV model, we extend the consistency of the weighted sum for the independent random variable sequence and obtain a result of convergence about the weighted sum for the exchangeable random variable sequence in EV model.

Key words

Errors-in-variables Model, The Weight Function Contains Real Variables, Convergence about the Weighted Sum.

1. Introduction

In the early 1930s, De Finetti put forward the concept of random variable exchangeability [1]. The so-called exchangeability of finite sequence $\{X_k\}_{k-1}^{n}$ refers to that if the joint distribution of random variables X_1, X_2, …, X_n is unchanged in displacement, that means the joint distributions to any displacement of 1,2,…,n on π, $X_{\pi(1)}$, $X_{\pi(2)}$,…, $X_{\pi(n)}$ and on X_1, X_2, …, X_n are identical. The infinite sequence $\{X_k\}_{k-1}^{+\infty}$ of exchangeable random variable is exchangeable, in case that any finite

subset thereof is exchangeable. Theories have proved that an infinite sequence of exchangeable random variables is independent identically distributed under the tail σ-algebraic condition, so no wonder that it is asymptotic similar to the independent identically distributed sequence.

As the fundamental structure theorem for infinite exchangeable random variable sequences, De Finetti's theorem states that an infinite exchangeable random variable sequence is independent and identically distributed with the condition of the tail σ-algebra. So, some results about independent identically distributed random variables are similar to exchangeable random variables. As the fundamental structure theorem for infinite exchangeable random variable sequences, De finetti's theorem is not applicable to finite exchangeable random variable sequences, and it is therefore necessary to find other technologies to solve the approximate behavior problems of finite exchangeable random variable sequences. By using the reverse martingale approach, scholars have given some results. In this paper, we do some researches about the similarity and differences between identically distributed random variable and exchangeable random variable sequences and mainly discuss the limit theory of exchangeable random variables.

We suppose X and Y are random variables and (X_i, Y_i), i=1, 2,.., n are the samples from the parent. Here the sample (X_i, Y_i) is a function of the importance of x.

$$W_{ni}(x) \triangleq W_{ni}(x, X_1, \cdots, X_n)$$

If the following conditions are satisfied:

$$(i) W_{ni}(x) > 0, i = 1, 2, \cdots, n;$$

$$(ii) \sum_{i=1}^{n} W_{ni}(x) = 1.$$

$W_{ni}(x)$ is called the weight function

The following questions are proposed in the literature [2]: Under what conditions, when n→∞, we obtain

$$\sum_{i=1}^{n} W_{ni}(x_0) Y_i \to E(Y | X = x) \quad a.s \tag{1}$$

Discussion on this issue is rare at home and abroad.

When we establish the regression model, we assume that the independent variable is fixed, and that the dependent variable is affected by random factors or measurement errors.

$$y = \mu + \varepsilon,\ \mu = f\left(x;\theta\right)$$

In the above equations, x and θ are vectors of dimension k, and y, μ, and ε are vectors of dimension n, and $E(\varepsilon)=0$.

For a long time, the research on EV model has not considered the errors in model selection; in other words, in the model, the dependent variable and the independent variable are really connected by the function, but later it was found that in economic analysis, there exist errors. In many cases in actual production, the independent variable is the measured value and thus will be affected by random factors. So when the parameter is a random variable, the model can be expressed as follows:

$$x = \xi + \delta,\ y = \mu + \varepsilon,\ \mu = f\left(\xi;\theta\right)$$

In the formula $E(\varepsilon)=E(\delta)=0$, ξ, and θ are represented as parameters, we call this model the errors-in-variables model.

There are some literatures about the errors-in-variables model [3-12], Generally we assume that $\tilde{X} = X + \varepsilon$ is right, when studying this model. Here X is a variable that cannot be directly observed while $\tilde{X}$ is a variable that can be observed. Under normal circumstances, the relationship between $\tilde{X}$ and X is complicated. For example, $\tilde{X} = \psi(X, \varepsilon)$, where ε is the error of measurement independent of (X, Y). ψ represents an arbitrary known function.

In the one-dimensional linear structural relation of the errors-in-variables model $y=a+bx$, $Y=y+\varepsilon$, $X=x+u$, if the parameters a,b are bounded continuous functions $a(t)$, $b(t)$ where the real variable $t \in (0,1)(b(t) \neq 0)$, then the EV model of the one-dimensional linear structure with variable coefficients is obtained as follows:

$$\begin{cases} y = a\left(t\right) + b\left(t\right)x \\ Y = y + \varepsilon,\ X = x + u \end{cases}$$

Where x, y are a random variables, and (ε,u) are measurement errors.

Suppose $t_0 \in (0,1)$, and we want to estimate the parameters $(a(t_0),b(t_0))$ at t_0. If we are not able to observe n times at t_0, we just have to observe n times in the vicinity of t_0. Supposing t_1, t_2,..., t_n, are n design points in [0,1], and satisfy $0 \leqslant t_1 < t_2 < \ldots < t_n \leqslant 1$, we observe (y,x) at every point t_i, and then will get n groups of observations of (Y_i, t_i, X_i), $i=1,2,\ldots n$. If we use these n groups of observations to estimate the parameters $(a(t_0),b(t_0))$ at t_0, we should note that the observed values of (Y_i, t_i, X_i) at t_i are not the same as those at t_0. The importance can be measured by the weight function $W_{ni}(t_0)$ of the real variable t_i, $i=1, 2, ..., n$. We first give the following definitions:

Suppose (Y_i, t_i, X_i), $i=1, 2, \ldots, n$. are the samples taken from the parent of (Y, X), that t_1, t_2, ..., t_n are n design points in [0,1], and that t_0 is a point within the interval (0,1). $W_{ni}(x) \triangleq W_{ni}(t_0, t_1, \cdots t_n)$ is the function of the real variables t_1, t_2, ..., t_n, ($i=1, 2, ..., n$), and we call it the real variable weight function, if it satisfies the following conditions:

$$(i) W_{ni}(t_0) > 0, i = 1, 2, \cdots, n;$$

$$(ii) \sum_{i=1}^{n} W_{ni}(t_0) = 1.$$

We assume the one-dimensional probability density function is $\kappa(\bullet)$ and that the bandwidth is $h_n \in (0,1/2)$, and then we obtain:

$$W_{ni}(t_0) = \frac{\kappa\left(\frac{t_i - t_0}{h_n}\right)}{\sum_{i=1}^{n} \kappa\left(\frac{t_i - t_0}{h_n}\right)}, i = 1, 2 \cdots, n$$

$W_{ni}(t_0)$ is called the kernel weight function.

2. Result

Here the weight function $W_{ni}(x)$ is a real variable kernel function. $W_{ni}(t_0)$ is studied in (1). We obtain the conclusion on the exchangeable random variables of $\{Y_i\}_{i=1}^{n}$, and obtain the consistency between the weighted sums of $\sum_{i=1}^{n} W_{ni}(t_0) Y_i$ and the weighted sum of the sequence of exchangeable random variables in the EV model.

The theorem assumes it satisfies the following conditions:

A_1 For any real variable kernel function $\{W_{ni}(t_0)\}_{i=1}^{n}$, there exists the integer A:

$$\max_{1\le i\le n} W_{ni}(t_0) \le \frac{A\log n}{nh_n}$$

A_2 Random variables of Y, Y_1, Y_2, …, Y_n are the exchangeable random variable sequence. $Cov(Y_1,Y_2)=0$, and EY exists, and there is a positive number D for which, $Var(Y)\le D$, so

(1) If $\frac{nh_n}{\log n}\to\infty(n\to\infty)$ holds, then $\sum_{i=1}^{n} W_{ni}(t_0)Y_i \xrightarrow{P} EY$

(2) If $\frac{\sqrt{n}h_n}{\log^2 n}\to\infty(n\to\infty)$ holds, then $\sum_{i=1}^{n} W_{ni}(t_0)Y_i \to EY$ *a.s.*

$\sum_{i=1}^{n} W_{ni}(t_0)Y_i \xrightarrow{P} EY$ and $\sum_{i=1}^{n} W_{ni}(t_0)Y_i \to EY$ *a.s.* can be described as follows:

$\sum_{i=1}^{n} W_{ni}(t_0)(Y_i - EY_i) \xrightarrow{P} 0$ and $\sum_{i=1}^{n} W_{ni}(t_0)(Y_i - EY_i) \to 0$ *a.s.*

$E(Y_i - EY_i)=0, Var(Y_i - EY_i)=Var(Y_i)\le D, i=1,2,\cdots,n,$; therefore, the theorem can be changed as follows:

Theorem 1 For the weight function $\{W_{ni}(t_0)\}_{i=1}^{n}$ of any real variable under the following conditions, there is a positive number A:

A_1 $\max_{1\le i\le n} W_{ni}(t_0) \le \frac{A\log n}{nh_n}$

A_2 Random variables of Y, Y_1, Y_2, …, Y_n are exchangeable random variables sequence. $Cov(Y_1,Y_2)=0$, and $EY=0$, and there is a positive number D, for which

$EY^2 = Var(Y) \le D$, so

(1) If $\frac{nh_n}{\log n}\to\infty(n\to\infty)$ holds, then $\sum_{i=1}^{n} W_{ni}(t_0)(Y_i - EY_i) \xrightarrow{P} 0$

(2) If $\frac{\sqrt{n}h_n}{\log^2 n}\to\infty(n\to\infty)$ holds, then $\sum_{i=1}^{n} W_{ni}(t_0)(Y_i - EY_i) \to 0$ *a.s.*

Proof: (1) because $E\left(\sum_{i=1}^{n} W_{ni}(t_0)Y_i\right)=0$, to prove $\sum_{i=1}^{n} W_{ni}(t_0)(Y_i - EY_i) \xrightarrow{P} 0$ is right, we only need to prove:

$$\lim_{n\to\infty} Var\left(\sum_{i=1}^{n} W_{ni}(t_0)Y_i\right) = 0$$

In fact, due to the result of A_1, A_2 and $\sum_{i=1}^{n} W_{ni}(t_0) = 1$, we obtain

$$Var\left(\sum_{i=1}^{n} W_{ni}(t_0)Y_i\right) = \sum_{i=1}^{n} W_{ni}(t_0)Var(Y_i) = D\left(\max_{1\le i\le n} W_{ni}(t_0)\right) \le \frac{DA\log n}{nh_n}$$

Because of the result of $\frac{nh_n}{\log n} \to \infty (n \to \infty)$, we know that $\lim_{n\to\infty} Var\left(\sum_{i=1}^{n} W_{ni}(t_0)Y_i\right) = 0$ is right, and then

$$\sum_{i=1}^{n} W_{ni}(t_0)(Y_i - EY_i) \xrightarrow{P} 0$$

(1) is proved.

(2) To prove $\sum_{i=1}^{n} W_{ni}(t_0)(Y_i - EY_i) \to 0$ *a.s.* is right, for any given positive number ε (suppose $\varepsilon<D/2$), we note

$$Y_i^{(1)} = Y_i I\left\{|Y_i| \le \varepsilon^2\sqrt{i}\right\}, Y_i^{(2)} = Y_i I\left\{|Y_i| > \varepsilon^2\sqrt{i}\right\}$$

Then we obtain $Y_i = Y_i^{(1)} + Y_i^{(2)}$, and based on $EY_i=0$, we obtain $EY_i^{(1)} = -EY_i^{(2)}$, so

$$Y_i = \left(Y_i^{(1)} - EY_i^{(1)}\right) + \left(Y_i^{(2)} - EY_i^{(2)}\right)$$

is right.

$$\sum_{i=1}^{n} W_{ni}(t_0)Y_i = \sum_{i=1}^{n} W_{ni}(t_0)\left(Y_i^{(1)} - EY_i^{(1)}\right) + \sum_{i=1}^{n} W_{ni}(t_0)\left(Y_i^{(2)} - EY_i^{(2)}\right) \quad (2)$$

There are two steps to prove the result. First we need to prove the result as follows: when $n\to\infty$,

$$\sum_{i=1}^{n} W_{ni}(t_0)\left(Y_i^{(1)} - EY_i^{(1)}\right) \to 0 \qquad a.s. \tag{3}$$

We note $Z_{ni} \triangleq W_{ni}(t_0)\left(Y_i^{(1)} - EY_i^{(1)}\right)$, and $b_n \triangleq \max_{1\le i\le n} W_{ni}(t_0)$, so $\{Z_{ni}\}_{i=1}^{n}$ is the zero mean exchangeable random variable sequence, and $b_n \le \dfrac{A\log n}{nh_n}$.

Because $\left|Y_i^{(1)} - EY_i^{(1)}\right| \le \left|Y_i^{(1)}\right| + \left|EY_i^{(1)}\right| \le \varepsilon^2\sqrt{i} + \varepsilon^2\sqrt{i} = 2\varepsilon^2\sqrt{i}$, so we obtain

$$\max_{1\le i\le n}\left|Z_{ni}\right| = \max_{1\le i\le n} W_{ni}(t_0)\left|Y_i^{(1)} - EY_i^{(1)}\right| \le \max_{1\le i\le n}\left|Y_i^{(1)} - EY_i^{(1)}\right| \le b_n \max_{1\le i\le n} 2\varepsilon^2\sqrt{i} = 2b_n\varepsilon^2\sqrt{n} < Db_n\varepsilon\sqrt{n}$$

Because

$$Var(Z_{ni}) = Var\left[W_{ni}(t_0)\left(Y_i^{(1)} - EY_i^{(1)}\right)\right] = W_{ni}^{\ 2}(t_0)\,Var\left(Y_i^{(1)}\right) \le \left(\max_{1\le i\le n} W_{ni}(t_0)\right)^2 Var(Y_i) \le b_n^{\ 2} D$$

So $\sum_{i=1}^{n} Var(Z_{ni}) \le nb_n^{\ 2} D$ is right.

Because $\dfrac{\log^2 n}{\sqrt{n}h_n} \to 0$ and $\dfrac{1}{\log n} \to 0$ are right when $n\to\infty$, for arbitrary ε>0, when n is sufficiently large, $\dfrac{\log^2 n}{\sqrt{n}h_n} < \varepsilon$ and $\dfrac{1}{\log n} < \varepsilon$ is right. From the result of Bennett exponential inequality, we note B=2DA(A+1), then

$$P\left\{\left|\sum_{i=1}^{n} Z_{ni}\right| \ge \varepsilon\right\} \le 2\exp\left\{-\frac{\varepsilon^2}{2nb_n^{\ 2}D + 2Db_n\varepsilon^2\sqrt{n}}\right\}$$

$$2\exp\left\{-\frac{\varepsilon^2}{\frac{B\varepsilon^2}{\log n}}\right\} = \frac{2}{n^{\frac{1}{B\varepsilon}}} < \frac{2}{n^2}\left(\varepsilon < \frac{1}{2B}\right)$$

Because $\sum_{i=1}^{n}\frac{1}{n^2}<\infty$ is right, $\sum_{i=1}^{n}P\left(\left|\sum_{i=1}^{n}Z_{ni}\right|\geq\varepsilon\right)<\infty$ is also right.

By Borel-Cantelli lemma, for arbitrary ε>0, when n is sufficiently large, we obtain:

$$\left|\sum_{i=1}^{n}Z_{ni}\right|<\varepsilon \qquad a.s.$$

When n is sufficiently large, we obtain $W_{ni}(t_0)\left(Y_i^{(1)}-EY_i^{(1)}\right)\to 0$ $\quad a.s.$ that is

$$\sum_{i=1}^{n}W_{ni}(t_0)\left(Y_i^{(1)}-EY_i^{(1)}\right)\to 0 \qquad a.s.$$

We prove the result as follows. When n is sufficiently large,

$$W_{ni}(t_0)\left(Y_i^{(2)}-EY_i^{(2)}\right)\to 0 \qquad a.s. \tag{4}$$

Firstly, $\left\{\left|Y_i^{(2)}-EY_i^{(2)}\right|\right\}_{i=1}^{n}$ is a zero mean exchangeable random variable sequence, so

$$E\left\{|Y_i|I\{Y_i\}\leq\varepsilon^2\sqrt{i}\right\}=\int_{-\infty}^{+\infty}|x|I\left\{|x|\leq\varepsilon^2\sqrt{i}\right\}dF_{Y_i}(x)\leq\varepsilon^2\sqrt{i}\int_{-\infty}^{+\infty}dF_{Y_i}(x)=\varepsilon^2\sqrt{i}$$

and so

$$\max_{1\leq i\leq n}E\left|Y_i^{(2)}\right|=\max_{1\leq i\leq n}E\left|Y_i^{(1)}\right|=\max_{1\leq i\leq n}E|Y_I|I\left\{|Y_I|\leq\varepsilon^2\sqrt{i}\right\}\leq\max_{1\leq i\leq n}\varepsilon^2\sqrt{i}=\varepsilon^2\sqrt{n}$$

$$\max_{1\leq i\leq n}\left|Y_i^{(2)}\right|-E\left|Y_i^{(2)}\right|\leq\max_{1\leq i\leq n}\left(\left|Y_i^{(2)}\right|+E\left|Y_i^{(2)}\right|\right)\leq 2\max_{1\leq i\leq n}\left|Y_i^{(2)}\right|\leq 2\varepsilon^2\sqrt{n}$$

Because

$$Var\left(\left|Y_i^{(2)}\right|-E\left|Y_i^{(2)}\right|\right)=E\left(\left|Y_i^{(2)}\right|-E\left|Y_i^{(2)}\right|\right)^2=Var(Y_i)\leq D$$

we have $\sum_{i=1}^{n} Var\left(\left|Y_i^{(2)}\right| - E\left|Y_i^{(2)}\right|\right) \le nD$.

Because

$$E\left\{\left|Y_i\right| I\left\{Y_i\right\} > \varepsilon^2\sqrt{i}\right\} = \int_{-\infty}^{+\infty} |x| I\left\{|x| > \varepsilon^2\sqrt{i}\right\} dF_{Y_i}(x) = \int_{|x|>\varepsilon^2\sqrt{i}} |x| dF_{Y_i}(x) = \varepsilon^2 i^{\frac{1}{2}} Var(Y_i) < \varepsilon^2 D i^{\frac{1}{2}}$$

when *n* is sufficiently large, we obtain

$$\sum_{i=1}^{n} E\left|Y_i^{(2)}\right| \le 2\sqrt{n}D\varepsilon^2 < \frac{1}{2}\varepsilon\sqrt{n}\log n$$

(we use the formula $\sum_{i=1}^{n} i^{\frac{1}{2}} < 2\sqrt{n}$ above. When *n* is sufficiently large, $4D < \varepsilon^2/\log n$.)

Because $\sum_{i=1}^{n} E\left|Y_i^{(2)}\right| \le 2\sqrt{n}D\varepsilon^2 < \frac{1}{2}\varepsilon\sqrt{n}\log n$ is right and due to the result of bennett exponential inequality, we obtain

$$P\left(\sum_{i=1}^{n}\left|Y_i^{(2)}\right| \ge \varepsilon\sqrt{n}\log n\right) \le 2\exp\left\{-\frac{\log n}{10\varepsilon}\right\} \le 2n^{-2} \quad \left(\varepsilon < \frac{1}{20}\right)$$

Because $\sum_{i=1}^{n} n^{-2} < \infty$ is right, we obtain

$$\sum_{i-1}^{n} P\left(\sum_{i=1}^{n}\left|Y_i^{(2)}\right| \ge \varepsilon\sqrt{n}\log n\right) < \infty$$

By Borel-Cantelli lemma, we obtain

$$\sum_{i=1}^{n}\left|Y_i^{(2)}\right| < \varepsilon\sqrt{n}\log n \qquad a.s$$

Because $\sum_{i=1}^{n} E\left|Y_i^{(2)}\right| \le 2\sqrt{n}D\varepsilon^2 < \frac{1}{2}\varepsilon\sqrt{n}\log n$ and $\sum_{i=1}^{n}\left|Y_i^{(2)}\right| < \varepsilon\sqrt{n}\log n$ *a.s.* is right, when n is sufficiently large, we have

$$\left|\sum_{i=1}^{n} W_{ni}(t_0)\left(Y_i^{(2)} - E\left|Y_i^{(2)}\right|\right)\right| = \frac{3}{2}\varepsilon A\frac{\log^2 n}{\sqrt{n}h_n} \quad a.s$$

Because of the result of $\frac{\log^2 n}{\sqrt{n}h_n} \to 0\left(n \to \infty\right)$, when n is sufficiently large,

$$\sum_{i=1}^{n} W_{ni}(t_0)\left(Y_i^{(2)} - E\left|Y_i^{(2)}\right|\right) \to 0 \quad a.s$$

(4) is proved.

When n is sufficiently large, because $\sum_{i=1}^{n} W_{ni}(t_0)Y_i \to 0$ *a.s.* and (2), (3), and (4) are all right, we obtain

$$\sum_{i=1}^{n} W_{ni}(t_0)\left(Y_i - EY_i\right) \to 0 \quad a.s$$

The theorem is proved.

Conclusion

In this paper, as the application of statistics, when the weight function uses real variables in errors-in-variables model, we extend the consistency of the weighted sum for the sequence of independent random variables, and obtain a result of convergence about the weighted sum for the sequence of exchangeable random variables in errors-in-variables model. The exchangeable random variables are independent and identically distributed so, it has some properties of independent identically distribution and is more applicable to some statistics. The errors-in-variables model is the real variable weight function. The kernel weight function of independent and identically distributed variables is generalized to the errors-in-variables model in the real variable weight function. This paper discusses the convergence of weighted sums of random variable.

References

1. B.D. Finetti, Funzione caratteristica di unfenomeno aleatorio, 1930, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat.Nat., no. 4, pp. 86-133.
2. R.F. Patterson, R.L. Taylor, Strong laws of large numbers for triangular arrays of exchangeable random variables, 1985, Stochastic Analysis and Applications, no. 3, pp. 171-187
3. G.A.F. Seber, Linear Regression Analysis, 1987, New York, Wiley.
4. P.X. Zhao, L.G. Xue, Variable selection for semiparametric varying coefficient partially linear errors-in-variables models, 2010, Journal of Multivariate Analysis, no. 8.
5. Y. Amemiya, Instrumental variable estimator for the nonlinear errors in variables model, 1985, Journal of Econometrics, no. 28, pp. 273-289.
6. J.H. You, G.M. Chen. Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model, 2005, Journal of Multivariate Analysis, vol. 97, no. 2, pp. 324-341.
7. A. Gut, Precise asympototics for record times and the associated counting process, 2002, Stoch Proc Appl, no. 101, pp. 233-239.
8. Q.H. Wang, Dimension reduction in partly linear error-in-response models with validation data, 2003, Journal of Multivariate Analysis, no. 85, pp. 234-252
9. O. Davidov, Estimating the slope in measurement error models a different perspective, 2005, Statistics & Probability Letters, vol. 71, no. 3, pp. 215-223
10. P.K. Shukla, M.E. Orazcm, O.D. Crisaue, Validation of the measurement model concept for error structure identification, 2004, Electrochemica Acta, no. 49, pp. 2881-2889.
11. H.Hong, E.Tamer, A simple estimator for nonlinear error in variable models, 2003, Journal of Econometrics, vol. 117, pp. 1-19.
12. H.F. Chen, J.M. Yang, Strong consistent coefficient estimate for errors-in-variables models, 2005, Automatica, no. 41, pp. 1025-1033.