Japan and the World Economy 36 (2015) 21–28

Contents lists available at ScienceDirect

Japan and the World Economy

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

A random walk stochastic volatility model for income inequality

Haruhisa Nishino [a,*], Kazuhiko Kakamu [b]

[a] Faculty of Law, Politics and Economics, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba-shi, Chiba 263-8522, Japan
[b] Graduate School of Business Administration, Kobe University, Japan

ARTICLE INFO

Article history:
Received 1 October 2014
Received in revised form 8 June 2015
Accepted 27 June 2015
Available online 16 July 2015

JEL classification:
C11
C22
D31

Keywords:
Income inequality
Lognormal
Persistence
Selected order statistics
Random-walk stochastic volatility (SV) model

ABSTRACT

This paper develops dynamic models that include income inequality from grouped income data to investigate persistent inequality. After we check that the lognormal distribution is adequately fitted to Japanese income data, by the asymptotic theorem of selected order statistics we construct an approximate linear model, which is extended to dynamic models, including a stochastic volatility (SV) model and a random-walk SV model. We can thus estimate the parameter of inequality directly. Both models are estimated using Japanese income data with the Markov chain Monte Carlo (MCMC) method and a model comparison is made. The SV model is better fitted than the random-walk SV model. We can capture the changing Gini coefficients for Japan using the SV model.

1. Introduction

Income inequality in a global economy has attracted a great deal of attention from many researchers, and it is a significant issue in the Japanese economy. Tachibanaki (2005) and Ohtake (2008) are famous in the field of income inequality in Japan. Moriguchi and Saez (2008) describe the historical change in income inequality in Japan with tax statistics. In particular, persistent inequality is an important problem in the field. Piketty (2000) provides a extensive survey of it. Lillard and Willis (1978), MaCurdy (1982), and Meghir and Pistaferri (2004) have made empirical studies of it using panel data analysis. Using a different method from them, this paper directly estimates the parameters of income inequality and examines the persistence of inequality in the same manner as Nishino et al. (2012), which developed a dynamic model including income inequality using a stochastic volatility (SV) model and pointed out that Japanese income inequality is persistent. Extending a model in Nishino et al. (2012), this paper attempts to determine whether the Japanese income inequality has a random-walk property or not.

In this paper, we use grouped data because microdata are difficult to access in Japan, and we use a parametric model in which we assume the lognormal distribution (Atoda et al. (1988) estimate income distributions in Japan using parametric distributions in a similar way). The lognormal distribution has been widely used to describe income since Gibrat (1931). We show in the next section that the lognormal distribution best fits the Japanese income data among of the two-parameter distributions including the gamma, Weibull and Pareto distributions.

Following the method in Nishino et al. (2012), we construct a dynamic model with the stochastic volatility (SV) model and estimate it with Markov chain Monte Carlo (MCMC) methods. Nishino et al. (2012) also showed that the SV model is better fitted to the Japanese income data than the model without SV. The result indicates the persistent income inequality, because the model without SV is a sequence of independent income distributions. In other words, it shows that the income equality has a time series structure represented with a stochastic volatility model. Moreover, the estimated autoregressive parameter of the stochastic volatility model was close to 1, which suggested the possibility of a random-walk SV model. If the random-walk model is supported, another type of persistent income inequality will be expected. For example, we expect that a shock to the volatility of the income would be permanent.

* Corresponding author. Tel.: +81 43 290 2422; fax: +81 43 290 2403.
E-mail address: nishino@le.chiba-u.ac.jp (H. Nishino).

http://dx.doi.org/10.1016/j.japwor.2015.06.003
0922-1425/

If it would be not supported, the empirical result would make the evidence of the SV model more robust. We are thus interested in extending the dynamic model with SV to a random-walk SV model in which a parameter representing inequality follows a random-walk process. We compare a SV model with a random-walk SV model for the Japanese income data.

This paper is organized as follows. Section 2 explains the lognormal distribution and its fit to the Japanese income data. It also describes an approximate linear model using grouped data. Section 3 explains the dynamic models that include the inequality parameter σ, which are an SV model and a random-walk SV model. Section 4 compares the two dynamic models for the Japanese income data (Family Income and Expenditure Survey from the Statistics Bureau of Japan's Ministry of Internal Affairs). Section 5 concludes the paper.

2. Lognormal distribution and the model

Since Gibrat (1931) introduced the lognormal distribution to representing income, it has become well known in the literature of income inequality. Recently, Battistin et al. (2009) wrote *"we confirm that the income distribution in countries including the United States and the United Kingdom has a shape that is close to, but not quite, log normal"*. Moreover the lognormal is convenient for analyzing inequality since a shape parameter σ of the lognormal only contains information on inequality. In this section, we show that the lognormal distribution is better than other two-parameter distributions for the Japanese income data. We next explain an approximate linear model based on selected order statistics that are obtained from group data.

2.1. Lognormal distribution and grouped income data

The lognormal distribution has the cumulative distribution function (cdf)

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \quad x > 0, \quad \sigma > 0 \tag{1}$$

and the probability density function (pdf)

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}. \tag{2}$$

The Gini coefficient of the lognormal distribution is

$$G = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1 \tag{3}$$

(see e.g., Kleiber and Kotz (2003)). Eq. (3) shows that a shape parameter σ only affects the Gini coefficient. That is, we are interested in estimating σ when analyzing inequality.

We next investigate whether the lognormal is better than other two-parameter distributions. For that purpose, we estimate the lognormal, gamma, Weibull and Pareto distributions by maximum likelihood estimation and compare them with the Akaike Information Criterion (AIC). For the estimation and the calculation of the AICs we need a likelihood for grouped data, which is obtained from a joint density function of selected order statistics in the following way.

Assume that the whole sample size is n and the subsample size is k. The subsample is ordered with selected order statistics $\{X(n_1) \leq X(n_2) \leq, \ldots, \leq X(n_k)\}$ for $1 \leq n_1 < n_2 < \ldots < n_k \leq n$. We let $(x_1, x_2, \ldots, x_k)$ denote observed values of selected order statistics $\{X(n_1), X(n_2), \ldots, X(n_k)\}$, then having a joint density $f(x_1, x_2, \ldots, x_k|\boldsymbol{\theta})$ of the selected order statistics (see e.g., David and Nagaraja, 2003, 2.2.2))

$$f(x_1, x_2, \ldots, x_k|\boldsymbol{\theta}) = n!\frac{F(x_i)^{n_1-1}}{(n_1-1)!}f(x_1) \times \left\{\prod_{i=2}^{k}\frac{(F(x_i)-F(x_{i-1}))^{n_i-n_{i-1}-1}}{(n_i-n_{i-1}-1)!}f(x_i)\right\}\frac{(1-F(x_k))^{n-n_k}}{(n-n_k)!}, \tag{4}$$

where θ is the parameter (μ, σ), $F(\cdot)$ is a cdf (1) and $f(\cdot)$ is a pdf (2). We thus have a likelihood function of θ

$$L(\boldsymbol{\theta}|x_1, x_2, \ldots, x_k) = f(x_1, x_2, \ldots, x_k|\boldsymbol{\theta}), \tag{5}$$

from the joint density $f(x_1, x_2, \ldots, x_k|\boldsymbol{\theta})$ of the selected order statistics.

Table 1
AIC

Year	Lognormal	Gamma	Weibull	Pareto
Workers' households				
1975	32188.22	32267.90	32415.14	32983.79
1985	32186.19	32246.75	32357.12	33064.03
1995	32185.36	32247.94	32357.87	33074.55
2005	32189.47	32243.20	32339.59	33139.38
Two-or-more-person households				
1975	32188.42	32265.67	32394.48	33024.64
1985	32193.70	32240.93	32324.25	33162.80
1995	32207.58	32238.79	32301.08	33269.65
2005	32184.50	32273.60	32391.89	32996.79

We explain the income data used in the paper, which is quintile data from the Family Income and Expenditure Survey published by the Statistics Bureau of Japan's Ministry of Internal Affairs. The income data contain two types households: workers' households and two-or-more person households. Total households include all types of households, which consist of two-or-more-person households and single person households. Two-or-more person households consist of workers' households and non-workers' fhouseholds including agricultural, forestry, and fishery households. Since pre-1995 data are available for two-or-more-person households and workers' fhouseholds, we use two-or-more-person household data and workers' fhousehold data in this paper.

We take an example of grouped data from the Family Income and Expenditure Survey (2005, workers' households, quintile data), where the whole sample size is n = 10, 000 and the subsample size is k = 4, and (n_1, n_2, n_3, n_4) = (2000, 4000, 6000, 8000). The example of grouped income data is

$$\left\{X(2000), X(4000), X(6000), X(8000)\right\} = (442, 582, 730, 944) \quad (\textit{ten} \quad \textit{thousand} \quad \textit{JPY}).$$

We compute AICs from the quintile income data using Ox version 7.0 (OS_X_64/U) (see Doornik, 2009). The whole sample size is n = 10, 000, and the subsample size is k = 4 owing to quintile data. Table 1 shows the AICs of four distributions (lognormal, gamma, Weibull, and Pareto) for 1975, 1985, 1995 and 2005.

Table 1 indicates that the AICs of the lognormal are smaller than those of the other distributions and that the lognormal is better fitted to the data than the other two-parameter distributions. We thus conclude that it is adequate to use the lognormal rather than other two-parameter distributions.

2.2. The approximate linear model based on selected order statistics

We here explain an approximate normal linear regression model using grouped data. Regarding grouped data as selected order