Economics Letters 120 (2013) 74–78

Contents lists available at SciVerse ScienceDirect

Economics Letters

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

How uncertain is household income in China

Jihai Yu, Guozhong Zhu *

Guanghua School of Management, Peking University, China

HIGHLIGHTS

- The stochastic processes of income in China are investigated based on longitudinal data from CHNS 1989–2009.
- Compared with the US households in PSID, income of the Chinese households is subject to greater uncertainty.
- Compared with the 1990s, income of the Chinese households has been subject to greater uncertainty in the 2000s.

ARTICLE INFO

Article history:
Received 14 December 2012
Received in revised form
3 March 2013
Accepted 6 March 2013
Available online 18 March 2013

JEL classification:
E21
E24
E25
P52

Keywords:
Income process
Income shocks
Wealth–income ratio

ABSTRACT

This paper studies the stochastic processes of household income in China using longitudinal data from CHNS 1989–2009. We consider both labor income and total household income. We find that (i) compared with the US households in PSID, income of the Chinese households is much more uncertain; (ii) compared with the 1990s, degree of income uncertainty in the 2000s is substantially higher in China.

1. Introduction

The stochastic property of household income is one of the key determinants of household decisions such as consumption–saving decision and portfolio choice.[1] In the context of Chinese economy, these decisions are linked to some prominent macroeconomic phenomena. For example, the income process is related to the unusually high savings rate which has become a puzzle to economists.[2] It is also related to the huge trade surplus of China that has caused tremendous disputes among its trading partners. Finally, it affects the optimal investment of the currently largest foreign reserve in the world, as the household finance literature has demonstrated that income risks shape the optimal portfolio to a large extent.

Despite the importance, little work has been done on the stochastic property of household income in China. In this paper we use longitudinal data from China Health and Nutrition Survey (CHNS) and minimum distance estimation to document the size and persistence of income shocks to the Chinese households. We consider both labor income and total household income. For comparison, income processes of US households during the same period of time are also estimated from the Panel Study of Income Dynamics (PSID).

We find that, income of the Chinese households is more uncertain than the US households. Breaking the CHNS data into two sub-periods (before and after 2000), we find that in the later period, labor income is subject to larger transitory shocks, while total household income is subject to larger persistent shocks. Overall, the degree of income uncertainty has increased substantially.

The rest of the paper is organized as follows. Section 2 introduces the data and basic statistics. Sections 3 and 4 discuss the econometric model and the estimation method. Section 5 reports the results.

* Corresponding author. Tel.: +86 10 62767407; fax: +86 10 62753624.
E-mail addresses: jihai.yu@gmail.com (J. Yu), gzhu@gsm.pku.edu.cn (G. Zhu).

1 For example, Carroll and Samwick (1997) and Carroll (2009) study consumption–saving problems based on the income process estimated from PSID. Bonaparte et al. (2012) study the link between income shocks and household portfolio choice.

2 See Modigliani and Cao (2004), Chamon and Prasad (2010) and Chao et al. (2011).

0165-1765/$ – see front matter
http://dx.doi.org/10.1016/j.econlet.2013.03.011

Table 1
Basic statistics.

	Number of households	Mean age	Std. dev. of age	Mean grade	Std. dev. of grade			
CHNS 1989–2009	452	36.7	8.1	6.1	5.7			
CHNS 1989–1997	1692	39.3	10.0	5.2	6.3			
CHNS 2000–2009	749	44.9	8.8	7.2	5.2			
PSID 1989–2009	1245	41.4	9.2	13.5	2.9			
	Average income in CHNS (in 2006 RMB)							
	1989	1991	1993	1997	2000	2004	2006	2009
Labor income	3856	3659	5276	11,014	12,830	14,975	21,099	34,489
Total income	4012	3877	5688	11,862	13,609	17,403	23,215	37,213
	Average income in PSID (in 2006 USD)							
	1989	1991	1993	1997	2001	2005	2007	2009
Labor income	39,237	41,741	44,706	48,265	59,337	62,897	67,173	69,267
Total income	42,979	45,561	49,436	55,081	70,689	72,337	79,098	82,130

The table reports the basic statistics of the CHNS and PSID samples used in the estimation. Income is all adjusted to the price level of 2006 based on CPI estimated by National Bureau of Statistics in China and Bureau of Labor Statistics in the US.

2. Data and basic statistics

We use the longitudinal data from the CHNS. Appendix A provides information about the data and sample selection criteria. Thus far, 8 waves of survey are available that span 20 years—1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009. We extract 3 panels from the data. The first one, called the 1989–2007 panel, includes households who satisfy sample selection criteria in each of the 8 waves of survey. The second one is called the 1989–1997 panel and includes households that satisfy the selection criteria in the first 4 waves of surveys. The third one is defined likewise and is named the 2000–2009 panel. We estimate and report income processes for each of the three panels.

The income processes of US households are estimated from the PSID. Corresponding to CHNS, the PSID sample is from 1989–2009. Before 1997, PSID survey was annual, since then it has been biannual.

Table 1 reports the basic statistics of both the CHNS and PSID samples used in our estimation.

3. The econometrics model

Guvenen (2009) reviews the literature of estimating stochastic income processes and characterizes two potential econometric models—the restricted income process (RIP) model and the heterogeneous income process (HIP) model. In the main text of this paper we report results from the HIP model which use less restrictive assumptions. Results from the RIP model are reported in Appendix C.

The following equations describe the income process of a typical household.

$$\log(Y_{i,a}) = f(X_{i,a}) + \hat{y}_{i,a} \quad a \in [1, \tau] \tag{1}$$

$$\hat{y}_{i,a} = \alpha_i + \beta_i a + z_{i,a} \tag{2}$$

$$z_{i,a} = p_{i,a} + \epsilon_{i,a} \tag{3}$$

$$p_{i,a} = \rho p_{i,a-1} + \eta_{i,a}. \tag{4}$$

In Eq. (1), $Y_{i,a}$ is the observed income of household i in age a, $X_{i,a}$ is a set of demographic variables associated with the deterministic component of income, and $\hat{y}_{i,a}$ is the income residual. In what follows, we use $a = 1$ to denote age 25 in the data when large fraction of the respondents have started their full-time job. Correspondingly, τ is the length of working life for a typical family in the data.

The variables α_i and β_i represent ex ante heterogeneity in income residual. Following Guvenen (2009), we assume a linear relation as captured by the term $\beta_i a$. The random vector (α, β) is distributed across individuals with zero mean, variances of σ_α^2 and σ_β^2, and covariance of $\sigma_{\alpha\beta}$.

The stochastic component of income, $z_{i,a}$, contains two types of income shocks: the persistent shock $\eta_{i,t}$ and the purely transitory shock $\epsilon_{i,t}$. The persistent shock follows an AR(1) process as shown in Eq. (4). The parameter ρ determines the persistence of shock $\eta_{i,t}$. Carroll and Samwick (1997) assume $\rho = 1$ and estimate the variances of shocks accordingly. Here we do not impose such a restriction.

We assume that both ϵ and η are normally distributed with zero mean and variances of σ_ϵ^2 and σ_η^2. These two variances, along with the persistence parameter ρ, capture the degree of income uncertainty. The correlation between ϵ and η is assumed to be zero.

4. Estimation method

We employ the minimum distance estimator to estimate parameters $\{\rho, \sigma_\eta^2, \sigma_\epsilon^2, \sigma_\alpha^2, \sigma_\beta^2, \sigma_{\alpha\beta}\}$ which characterize the heterogeneous income process. Specifically, we calculate a set of data moments that capture data features—variances and covariances of income residual $\hat{y}_{i,a}$ that is obtained from regressing $\log(Y_{i,t})$ on $X_{i,t}$.[3] We then derive the analytical expression of the corresponding model moments based on Eqs. (1)–(4). Since the model moments are functions of parameters, we can search for the set of $\{\rho, \sigma_\eta^2, \sigma_\epsilon^2, \sigma_\alpha^2, \sigma_\beta^2, \sigma_{\alpha\beta}\}$ that minimizes the distance between model moments and data moments.

Omitting household subscript i, we derive the following expressions from Eqs. (1)–(4).

$$\begin{aligned} \mathrm{var}(\hat{y}_a) &= \rho^{2a}\mathrm{var}(z_0) + \frac{1-\rho^{2a}}{1-\rho^2}\sigma_\eta^2 \\ &\quad + \sigma_\epsilon^2 + \left[\sigma_\alpha^2 + 2\sigma_{\alpha\beta}a + \sigma_\beta^2 a^2\right] \end{aligned} \tag{5}$$

$$\begin{aligned} \mathrm{cov}(\hat{y}_a, \hat{y}_{a-j}) &= \rho^{a(a-j)}\mathrm{var}(z_0) + \rho^j \frac{1-\rho^{2(a-j)}}{1-\rho^2} \\ &\quad \times \sigma_\eta^2 + \sigma_\alpha^2 + \sigma_{\alpha\beta}(2a-j) + \sigma_\beta^2 a(a-j) \end{aligned} \tag{6}$$

for $a \in [1, \tau]$ and $j \in [1, a]$. Here $\mathrm{var}(z_0)$ is the initial variance of household income. In practice, it is approximated by the variance of income residual of households whose heads are 25 years of age or younger.

Given a panel of households, the number of variances and covariances depends on number of cross sections in the available data. Let the number of cross sections be T, then for each age group, the number of moments is $\frac{T(T+1)}{2}$. For example, if we use CHNS

[3] We include in $X_{i,t}$ cubic polynomials of age and dummies of education attainment, occupation, sector of employment of the family head, urban–rural status, region of residence and year of survey.