



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.

© 2013 Elsevier B.V. All rights reserved.

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), gzhugsm.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).

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.

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

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

$$z_{i,a} = p_{i,a} + \epsilon_{i,a} \quad (3)$$

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

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).

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

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

for $a \in [1, \tau]$ and $j \in [1, a]$. Here $\text{var}(z_0)$ 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.

Table 2
Income processes in China and the US.

	var(ϵ) (1)	var(η) (2)	ρ (3)	var(α) (4)	var(β) (5)	cov(α , β) (6)	W/Y (7)	W/Y* (8)
Labor income (CN)	0.424 (0.076)	0.065 (0.053)	0.792 (0.168)	4.14E–12 (0.298)	0.00020 (0.00058)	–0.0011 (0.0128)	5.390	2.991
Total income (CN)	0.368 (0.062)	0.063 (0.051)	0.783 (0.203)	2.97E–10 (0.311)	0.00014 (0.00056)	3.34E–05 (0.0129)	4.942	2.981
Labor income (US)	0.035 (0.009)	0.084 (0.011)	0.690 (0.047)	0.211 (0.064)	0.00051 (0.00014)	–0.0066 (0.003)	2.756	2.756
Total income (US)	0.054 (0.004)	0.035 (0.004)	0.840 (0.036)	0.035 (0.073)	0.00012 (0.00012)	0.0012 (0.0028)	2.357	2.357

The table reports the parameters of income processes and the wealth–income ratios implied by the stochastic component of income in a standard precautionary saving model. W/Y* is the ratio from imposing var(ϵ) from the US data to both countries. In parenthesis are standard errors.

from 1989–2009, we have 8 wave of surveys in total. Therefore, for each age group, 36 covariances are available. Among them there are 8 variances and 28 covariances. Correspondingly, model moments are the averages of the variances and covariances over different age groups.

5. Results

5.1. Estimation results

Table 2 reports the results from both CHNS and PSID. Comparing the income processes in China to those in the US, we find that transitory shocks are about 6–12 times larger in China.⁴ Regarding persistent shocks, those from CHNS have a smaller variance but higher level of persistence in terms of labor income. In terms of total income, they have a larger variance but lower level of persistence. Overall, persistent shocks in China imply higher degree of income uncertainty, which we show in the next subsection.

Consistent with Guvenen (2009), based on our PSID sample of labor income, variance of income growth (i.e., var(β)) is statistically significant. However, it is much smaller and statistically insignificant in CHNS. Therefore, for the Chinese households, larger fraction of fanning out in income profiles is attributable to random shocks rather than pre-determined heterogeneity. For total household income in CHNS var(β) is not significant either.

Based on the same data we also estimate the RIP model. Results are reported in Appendix C. From the RIP model, households in China are also faced with greater income uncertainty than the US households. Comparing the two subperiods, we find that total household income in China has become more uncertain, but labor income has become less uncertain due to the decrease of the variance of persistent shocks.

5.2. Wealth–income ratio

It is well understood in the precautionary saving literature that greater income uncertainty leads to higher wealth–income ratio. Conversely, ceteris paribus, higher wealth–income ratio implies great income uncertainty. To better understand how uncertain the income of Chinese households is, we feed the estimated income processes into a standard precautionary saving model, and compare the implied wealth–income ratios. The following optimization problem is solved.

$$\max_{b_{t+1}, c_t} E \sum_{t=0}^{\infty} \delta^t \frac{c_t^{1-\gamma}}{1-\gamma}$$

s.t.

$$c_t = y_t + (1+r)b_t - b_{t+1}$$

where c , b and y stand for consumption, saving and income level of a household. We use standard parameters, $\delta = 0.96$ and $\gamma = 2$, to compute the implied average wealth–income ratios. Since we intend to understand the income uncertainty implied in the stochastic income processes, we assume $\beta = 0$ for each household so that the difference in wealth–income ratios reflects difference in size and persistence of income shocks.⁵ Results are reported in column (7) of Table 2. The implied wealth–income ratios in China is about twice of those in the US. To further understand how much of the difference is caused by persistent shocks, we restrict variance of transitory shocks in China to be the same as that in the US, and solve the model again. The results are reported in column (8) of Table 2. Still, the wealth–income ratios are higher in China. Thus, looking at the persistent shocks only, income in China is also more uncertain.

5.3. Comparison between 1989–1997 and 2000–2009

To understand how the stochastic income processes evolve in China, we break the CHNS sample into two sub-periods: 1989–1997 and 2000–2009, and report estimation results in Table 3. The variance of transitory shocks has increased for both labor income and total income. Persistent shocks have a smaller variance but increased level of persistence for labor income, but a substantially larger variance and slightly decreased persistence for total income. As shown in column (7), implied wealth–income ratio has become 26% higher based on labor income, and 8.2% higher based on total income. Thus in the later sub-period, households are subject to significantly greater income uncertainty. In terms of persistent shocks alone, the degree of uncertainty has also increased, as is evident in column (8).

6. Conclusion

We use the longitudinal data from CHNS 1989–2009 to estimate the stochastic processes of both labor income and total household income in China. Compared with the income processes in the US based on PSID, both types of income in China are subject to much higher degrees of uncertainty. Breaking the sample into two sub-periods, we find that both types of income in China are more uncertain in the 2000s relative to the 1990s.

Acknowledgments

We thank the editor Pierre-Daniel Sarte and one anonymous referee for their comments and suggestions for improving this paper. We also appreciate the research assistance from Kunyuan Qiao. Yu acknowledges funding from National Science Foundation of China (Grant No. 71171005).

⁴ Part of the difference in var(ϵ) might be caused by measurement errors.

⁵ We report the implied wealth–income ratios with heterogeneous β , and discuss how to compute the ratios in Appendix D.

Table 3
Income processes in the 1990s and 2000s.

	var(ϵ) (1)	var(η) (2)	ρ (3)	var(α) (4)	var(β) (5)	cov(α, β) (6)	W/Y (7)	W/Y* (8)
Labor income (1990s)	0.336 (0.069)	0.096 (0.062)	0.790 (0.2)	0.025 (0.868)	0.0010 (0.0022)	−0.0122 (0.0468)	5.443	5.443
Labor income (2000s)	0.437 (0.108)	0.063 (0.073)	0.887 (0.376)	0.305 (3.657)	0.0003 (0.0043)	−0.0114 (0.1173)	6.847	6.131
Total income (1990s)	0.324 (0.061)	0.073 (0.046)	0.779 (0.189)	3.56E−09 (0.782)	4.66E−05 (0.0018)	0.0015 (0.0399)	4.834	4.834
Total income (2000s)	0.326 (0.192)	0.116 (0.168)	0.736 (0.472)	0.500 (2.777)	0.0003 (0.0046)	−0.0132 (0.1177)	5.229	5.216

The table reports the parameters of income processes in China based on CHNS 1989–1997 and CHNS 2000–2009 and the wealth–income ratios implied by the stochastic component of income in a standard precautionary saving model. W/Y* is the ratio from imposing var(ϵ) from 1990s to both periods. In parenthesis are standard errors.

Table B
Comparison of the basic statistics between the selected sample and larger sample.

	Average income in CHNS (in 2006 RMB)							
	1989	1991	1993	1997	2000	2004	2006	2009
Labor income (selected sample)	3856	3659	5276	11,014	12,830	14,975	21,099	34,489
(Larger sample)	3861	3767	4908	8,678	10,706	12,220	14,749	25,631
Total income (selected sample)	4012	3877	5688	11,862	13,609	17,403	23,215	37,213
(Larger sample)	4306	4322	5885	10,516	12,755	16,845	18,994	30,586
	Average income in PSID (in 2006 USD)							
	1989	1991	1993	1997	2001	2005	2007	2009
Labor income (selected sample)	39,237	41,741	44,706	48,265	59,337	62,897	67,173	69,267
(Larger sample)	35,568	37,185	40,104	41,490	50,807	54,034	56,389	56,863
Total income (selected sample)	42,979	45,561	49,436	55,081	70,689	72,337	79,098	82,130
(Larger sample)	40,155	42,093	45,606	49,789	64,089	68,678	70,878	72,621

The table reports basic statistics of the larger samples that have minimal criteria in selection. For comparison, it also reports basic statistics of the selected sample used in the estimation of income processes.

Appendix A. Data appendix

CHNS. China Health and Nutrition Survey (CHNS) is an ongoing international collaborative project. The CHNS conducts surveys in regions that vary substantially in terms of geography, economic development, public resources, and health indicators over a 3-day period. CHNS uses a multistage, random cluster process to draw a sample of about 4400 households with a total of 26,000 individuals for each wave.⁶ The first wave of survey was conducted in 1989, followed by 1991, 1993, 1997, 2000, 2004, 2006, 2009 and 2011 waves during which surveyed households were revisited. We exclude the 2011 survey since they are still being proofread. The survey fits our study well because it collects detailed data on nine categories of income for each household—business, farming, fishing, gardening, livestock, non-retirement wages, retirement income, subsidies, and other income. In addition, CHNS also provides a rich set of demographic variables of household members.

Sample selection. We apply exactly the same sample selection criteria to both CHNS and PSID, so that income processes from the two data sets are comparable. Specifically, we select households that have valid information on income, rural–urban status, region of residence, as well as the following information for household heads: age, gender, education attainment, occupation and sector of employment. The following households are excluded: (i) households with zero income; (ii) households whose income grow by more than 2000% between any two surveys; (iii) households whose income drop by more than 2000% between any two surveys. Finally, we drop the low income SOE sample from PSID.

Our sample selection criteria leave us with households whose income is positive and not extremely volatile, the resulting income

Table C
Restricted income processes.

	var(ϵ)	var(η)	ρ
Comparison between China and the US			
Labor income (CN 1989–2009)	0.412 (0.042)	0.064 (0.021)	0.864 (0.024)
Total income (CN 1989–2009)	0.367 (0.027)	0.051 (0.013)	0.886 (0.016)
Labor income (US 1989–2009)	0.099 (0.007)	0.029 (0.003)	0.946 (0.004)
Total income (US 1989–2009)	0.068 (0.005)	0.022 (0.002)	0.950 (0.003)
Comparison between the 1990s and 2000s			
Labor income (CN 1989–1997)	0.362 (0.031)	0.080 (0.026)	0.852 (0.028)
Labor income (CN 2000–2009)	0.479 (0.087)	0.041 (0.039)	0.888 (0.05)
Total income (CN 1989–1997)	0.351 (0.023)	0.049 (0.013)	0.891 (0.017)
Total income (CN 2000–2009)	0.421 (0.094)	0.044 (0.044)	0.879 (0.059)

The table reports the parameters of stochastic income processes estimated from CHNS and PSID based on the RIP model. In parenthesis are standard errors.

is higher than the full samples from CHNS and PSID. We report the means of both types of income from various waves in [Appendix B](#). In addition, the means of age and grade in the selected sample are also different. Full sample of CHNS 1989–2009 has mean age of 46.3 and mean grade of 6.0, while the numbers in the selected sample are 36.7 and 6.1. Since we control for income level, grade and age when obtaining income residuals, these differences would not bias the estimation of stochastic income processes if the nature of income shocks does not depend on income level or age or educational attainment. In the literature there exists some evidence that more educated individuals are subject to less

⁶ More details are available in <http://www.cpc.unc.edu/projects/china/>.

Table D

Wealth–income ratio from the fully-fledged HIP model.

	var(ϵ) (1)	var(η) (2)	ρ (3)	var(α) (4)	var(β) (5)	cov(α, β) (6)	W/Y ratio (7)
Labor income (CN)	0.424 (0.076)	0.065 (0.053)	0.792 (0.168)	4.14E–12 (0.298)	0.00020 (0.00058)	–0.0011 (0.0128)	5.699
Total income (CN)	0.368 (0.062)	0.063 (0.051)	0.783 (0.203)	2.97E–10 (0.311)	0.00014 (0.00056)	3.34E–05 (0.0129)	5.135
Labor income (US)	0.035 (0.009)	0.084 (0.011)	0.690 (0.047)	0.211 (0.064)	0.00051 (0.00014)	–0.0066 (0.003)	3.478
Total income (US)	0.054 (0.004)	0.035 (0.004)	0.840 (0.036)	0.035 (0.073)	0.00012 (0.00012)	0.0012 (0.0028)	2.703

The table reports wealth–income ratios implied by the estimated heterogeneous income processes in China (CHNS) and the US (PSID).

income uncertainty. Given that, we might have understated the size and persistence of income shocks in both countries.

Appendix B. Basic statistics in the selected samples versus larger samples

As an effort to reduce measurement error, we exclude households whose income grows or drops by more than 2000% between any two surveys. Households with zero income at any time are also excluded. In addition, we exclude households that do not have valid information on age, educational attainment, occupation, sector of employment, region of residence and urban–rural status. Hence the selected sample is much smaller than the population. For comparison we calculate basic statistics from a much less selective sample, one that only requires valid information on age, grade and income. Results are reported in rows labeled “larger sample” in Table B. Compared with the more selective samples used for estimation, larger samples generally have lower income because they include zero income households. Income in the larger sample of CHNS has lower growth rate than that from selected sample, reflecting the increasing unemployment rate due to the reform of state-owned enterprises.

Appendix C. Results from the restricted income process (RIP) model

Much of the calibrated macroeconomic model uses the RIP model to approximate stochastic income process. Using the same procedure as described in the main text, we estimate the RIP processes of households in CHNS and PSID. Table C reports the results.

Appendix D. Wealth–income ratios implied by the fully-fledged HIP model

In the main text we report the wealth–income ratios from the HIP parameters, assuming $\beta = 0$ for each household. Here we

discussion the wealth–income ratios implied by a fully-fledged HIP model with heterogeneous α and β .

Recall that α determines ex ante heterogeneity in the income level. When the preference is homogeneous, income level does not affect wealth–income ratio, hence it is adequate to deal with the income growth rate β .

Since β enters the logarithm of income additively, it enters income level in a multiplicative way. It is well-understood that the household optimization problem with multiplicative component of income can be re-scaled. In our case, an individual faced with interest rate r and income growth rate g should have the same wealth–income ratio as an individual faced with interest rate $(1+r)/(1+g) - 1$ and zero income growth. Therefore in the computation, we first draw 100 income growth rates from the distribution determined by $\text{var}(\beta)$. For each income growth rate g , we transform the problem so that income growth rate is zero, but interest rate is $(1+r)/(1+g) - 1$. Now this is a standard optimization problem as under RIP parameterizations. We discretize the income process, compute the policy function and simulate the economy to obtain the wealth–income ratio implied by each particular income growth rate. Then we calculate the average wealth–income ratio of individuals with different income growth rates. Table D reports the ratios from various heterogeneous income processes.

References

- Bonaparte, Y., Cooper, R., Zhu, G., 2012. Consumption smoothing and portfolio rebalancing: the effects of adjustment costs. *Journal of Monetary Economics* 59, 751–768.
- Carroll, C.D., 2009. Precautionary saving and the marginal propensity to consume out of permanent income. *Journal of Monetary Economics* 56, 780–790.
- Carroll, C.D., Samwick, A.A., 1997. The nature of precautionary wealth. *Journal of Monetary Economics* 40, 41–71.
- Chamon, M.D., Prasad, E.S., 2010. Why are saving rates of urban households in China rising? *American Economic Journal: Macroeconomics* 2 (1), 93–130.
- Chao, C.-C., Laffargue, J.-P., Yu, E., 2011. The Chinese saving puzzle and the life-cycle hypothesis: a revaluation. *China Economic Review* 22, 108–120.
- Güvenen, F., 2009. An empirical investigation of labor income process. *Review of Economic Dynamics* 12, 58–79.
- Modigliani, F., Cao, S.L., 2004. The Chinese saving puzzle and the life-cycle hypothesis. *Journal of Economic Literature* 42, 145–170.