

# Co-integration Analysis of Rural Labor Non-agricultural Activities and Household Income

Fan SHI <sup>1,a</sup> Xue-Hui ZHANG <sup>2</sup>, and Shu-Wei CHEN <sup>3,b</sup>

<sup>1</sup>*Economics and Management School, Wuhan University, Wuhan, China 43007*

<sup>2</sup>*Dongguan Securities Research Institute, Shenzhen, China 518000*

<sup>3</sup>*School of Public Administration, Henan University of Finance and Economics, Zhengzhou, China 450046*

<sup>a</sup>*afanti525@hotmail.com*, <sup>b</sup>*shuweichen@126.com*

**Abstract:** The role of non-agricultural activities of rural labors in their household income is a noteworthy topic. This paper focuses on the long-term and stable equilibrium relationship between the two. Based on 1985 - 2012 Statistical Yearbook data, this study conducted corresponding calculation reasoning and co-integration test of the non-agricultural activities of rural labors and their household income, with the results showing: Long-term equilibrium relationship exists between the two, witnessing short-term fluctuations and a trend of short-term fluctuations to long-run equilibrium.

## 1 Introduction

Labor migration is an important way of rural residents to improve living standards, obtain fair opportunities to compete and transfer from low income to high income. <sup>[1]</sup>With the migration of rural labors to urban non-agricultural sectors increasingly frequently and regularly, rapid growth of their wages and consequent constant growth of their household income, the rural household income inequality appears. Relationship between the migration of rural labors to urban non-agricultural sectors and their household income and its impact on their household income inequality become noteworthy.

Multiple scholars have made academic researches on the impact of labor migration on the income inequality. Some believe labor migration helps reduce or mitigate income inequality. For example, Li Shi (in 1999)<sup>[2]</sup> proposed that labor migration could improve their household income, and contribute to inhibiting their income inequality. Cai Fang and Wang Meiyun (in 2009)<sup>[3]</sup> suggested that the increasing labor migration and growing income inequality structured a paradox due to ignorance or underestimation of the income “resident migrants”, namely migratory groups, resulting in underestimation of rural income and exaggeration of urban income. Others believed labor migration might expand rural household income inequality. For example, Yan Haokun and Xu Minghui (in 2008)<sup>[4]</sup> stated that against the household registration system, namely an “access threshold”, in other words, institutional constraints, labor migration expanded instead of reduced income inequality. Zhu Changcun and Ma Jingzhi (in

2009)<sup>[5]</sup> commented that rural human capital characterized by inner overflow only made competitive human capital workforce obtain high income, making labor migration expand income inequality to certain extent. However, some scholars argued that it was improper to simply conclude labor migration would expand or narrow the income inequality, which should be viewed separately depending on the circumstances. For instance, Liao Xianlang (in 2012)<sup>[6]</sup> analyzed the reason for synchronous expansion of labor migration and income inequality and suggested labor migration narrowed the income inequality without reflecting its effect against labor migration restrictions and unbalanced regional economic development. Cai Wu and Cheng Xiaojun (in 2012)<sup>[7]</sup> thought that labor migration and income inequality witnessed benign interaction and simultaneous reduction of relative income of disadvantaged groups, exacerbating the predicament of impoverished poor and low-income groups.

This paper conducts an overall analysis of the relationship between labor migration and income instead of any value judgment. It starts with labor migration, analyzes the relationship between labor migration income effects and household income and the role in rural household income with the co-integration method.

## 2 Variable Selection, Data Description and Research Methods

### 2.1 Variable Selection and Data Description

This paper takes rural household annual per capita net

income (RHAPCNI) as the dependent variable, the number of non-agricultural employed rural labors (NNENNERL) and rural household per capita wage income (RHPCWI) as the independent variable. It takes the time series data of 28 year (1985 to 2012) as the sample observations. Based on the data continuity and integrity, it has the data from previous *China Statistical Yearbooks* and takes 1985 as the base year, smoothing out the price fluctuation and eliminating the influence of the price factors.

It is necessary to explain **the measurement of the number of non-agricultural employed rural labors**. In terms of the definition of the number of non-agricultural employed rural labors, the total amount of migrant workers, out-of-town migrant workers and local migrant workers has been announced since 2007, lacking systematic statistics before 2007, sufficient samples, continuity and measurement analysis, and therefore it is necessary to find alternative data. China Statistical Yearbooks indicate the number of employed rural residents and the primary industry employees. The number of urban labors engaged in the primary industry is extremely limited, even negligible. Therefore, this paper

takes the difference between the number of employed rural residents and the number of the primary industry employees as the alternative measurement standard for urban labors engaged in non-agricultural secondary and tertiary industries. **In terms of measurement income variables of migrant workers**, based on the consistency with per capita household income variable, this paper takes per capita income of rural household as an alternative, and consequently per capita wage income of rural households as an alternative to the income of migrant workers.

In addition, logarithm taken by the dependent variable and the independent variable can eliminate time-series heteroscedasticity. At the same time, logarithmic transformation of data does not change the original co-integration relationship, achieving a linear trend. In data processing, this paper takes the natural logarithm of the independent variable and the dependent variable respectively: logarithm of rural household per capita wage income (LRHAPCNI), logarithm of the number of non-agricultural employed rural labors (LNNERL) and logarithm of rural household per capita wage income (LRHPCWI), as shown in Table 1:

**Table 1.** RHAPCNI, NNERL, RHPCWI and natural logarithm of China's rural residents from 1985 to 2012

Year	RHAPCNI	NNERL	RHPCWI	LRHAPCNI	LNNERL	LRHPCWI
1985	397.6	5935	72.2	5.985	8.689	4.279
1986	399.4	6735	76.9	5.990	8.815	4.343
1987	410.4	7337	84.7	6.017	8.901	4.439
1988	411.6	7818	89.0	6.020	8.964	4.489
1989	380.9	7714	86.4	5.943	8.951	4.459
1990	415.7	8794	84.1	6.030	9.082	4.432
1991	419.5	8928	89.9	6.039	9.097	4.499
1992	443.4	9592	104.3	6.094	9.169	4.647
1993	458.5	10866	96.8	6.128	9.293	4.573
1994	492.3	12174	106.0	6.199	9.407	4.663
1995	541.4	13495	121.4	6.294	9.510	4.799
1996	612.6	14208	143.4	6.418	9.562	4.966
1997	648.5	14199	159.7	6.475	9.561	5.073
1998	677.5	13844	179.7	6.518	9.536	5.191
1999	703.3	13214	200.5	6.556	9.489	5.301
2000	717.6	12892	223.7	6.576	9.464	5.410
2001	747.7	12276	243.9	6.617	9.415	5.497
2002	785.4	11481	266.6	6.666	9.348	5.586
2003	818.9	11302	286.8	6.708	9.333	5.659
2004	875.0	12141	297.5	6.774	9.404	5.695
2005	949.0	12816	342.4	6.855	9.458	5.836
2006	1030.5	13407	394.9	6.938	9.504	5.979
2007	1128.5	13637	435.1	7.029	9.521	6.076
2008	1218.5	13538	474.5	7.105	9.513	6.162
2009	1323.0	13616	529.2	7.188	9.519	6.271
2010	1466.9	13487	602.5	7.291	9.509	6.401
2011	1634.4	13912	694.2	7.399	9.541	6.543
2012	1809.1	13829	787.8	7.501	9.535	6.669

Source: Calculated according to China Statistical Yearbooks

## 2.2 Research Methods

This paper analyzes the relationship between income of non-agricultural employment of rural labors, migrant income and rural household income. According to analysis needs, the following regression model is created:

$$RHAPCNI_t = \beta_0 + \beta_1 NNERL_t + \beta_2 RHPCWI_t + \mu_t \quad (1)$$

In Formula (1), RHAPCNI represents rural household per capita income, NNERL represents the number of non-agricultural employed rural labors, RHPCWI represents migrant income replaced by per capita wage income,  $\mu$  represents random error to contain variables the formula does not take into account or unobservable

factors. In order to eliminate lead variance which may exist to conduct quantitative logarithmic conversion at both sides of Formula (1) without altering original time series co-integration relationship, expressed as follow:

$$\text{LRHAPCNI}_t = C + \beta_1 \text{LNNERL}_t + \beta_2 \text{LRHPCWI}_t + \varepsilon_t \quad (2)$$

In Formula (2),  $C$  is a constant term.

Specific analysis ideas are as follows:

Firstly, stability test, namely, unit root test of variables, was conducted. This study mainly adopted ADF test.

Secondly, based on the stability test, co-integration test was conducted for long-term and stable equilibrium between variables.

Thirdly, co-integration test is a long-term, balanced and stable relationship test, flawed to short-term equilibrium relationship. Error correction model (ECM) provides a way to examine the co-integrated short-term behavior so as to integrate the long-term relationship and short-term relationship between the variables. Thus, it is necessary to repeat the error correction model analysis.

In the end, Granger causality test was conducted,

which is the final variable regression verification.

### 3 Empirical Test and Result Analysis

Co-integration theory is a quantitative analysis method developed on the basis of time series vector auto-regression analysis outweighing traditional linear regression in terms of statistical stringency and scientificity.

#### 3.1 Variable Stability Test (ADF Test)

In case of time series quantitative analysis, if two or more data are not stable, the regression may be spurious. Normally, spurious regression may be caused by unstable time series, making the regression results distorted<sup>[8]</sup>. Thus, in case of regression analysis of the time series model, it is necessary to conduct the variable stability test. EVIEWS7.2 was used to conduct ADF test of the original series and first difference analysis. Specific test results are shown in Table 2

**Table 2.** Stability (ADF) test results

Series	ADF test			
	Inspection type (C, T, L)	ADF statistics	Threshold	Conclusion
LRHAPCNI	(C,T,1)	-1.3951	-3.2335*	Unstable
LNNERL	(C,0,1)	-2.0300	-2.6299*	Unstable
LRHPCWI	(C,T,1)	-1.5639	-3.2292*	Unstable
$\Delta$ LRHAPCNI	(C,0,1)	-4.2924	-2.9810**	Stable
$\Delta$ LNNERL	(0,0,1)	-2.0809	-1.9550**	Stable
$\Delta$ LRHPCWI	(C,0,1)	-3.714	-3.7115***	Stable

Note:  $\Delta$  represents first order difference; In (C, T, L), C, T, L represents respectively intercept term, trend term and lag order; ADF test lag order is determined by AIC rule, \*\*\*, \*\*, \* respectively represent significant drop against 1%, 5% and 10%.

The test results shown in Table 2 indicate the original series LRHAPCNI, LNNERL and LRHPCWI surpass the threshold 10% significance level, that is, series value presents unstable characteristics. However, after first order difference, they have their series converted to  $\Delta$ LRHAPCNI,  $\Delta$ LNNERL and  $\Delta$ LRHPCWI. Their ADF test results show that their statistics are less than 5%, 5% and 1% significance threshold, ADF test after difference is stable, namely first order integration, meeting the co-integration test prerequisite.

#### 3.2 Co-integration Test

Although the difference method can eliminate the series instability trend, the converted series often limits the economic issues discussed, even loses its economic significance. To eliminate the above effects, Engle & Granger (in 1987) suggested that the series itself was instable, but its linear combination might be a stable

series. The co-integration equation could explain the long-term stable equilibrium relationship between series.

In terms of co-integration equation long-term and stable equilibrium relationship, it is proper to analyze based on 2 test objects: one is co-integration test based on regression coefficients, and this study adopted OLS for regression estimation. The other is co-integration test based on regression residuals, and this study adopted ADF test. As it relates to 3 variables, it is necessary to conduct the co-integration test with the 2 independent variables and the dependent variable respectively. If both are successful, the co-integration equation witnesses a stable equilibrium relationship. The specific test is as follows:

**Firstly, co-integration test between variables LRHAPCNI and LNNERL:**

The first step, use OLS to estimate co-integration equation of the 2 variables:

$$\begin{aligned} \text{LRHAPCNI} &= -6.2742 + 1.3751\text{LNNERL} & (3) \\ & (-2.7670) \quad (5.6569) \\ R^2 &= 0.8517 \quad \text{DW} = 0.0712 \quad F = 32.0001 \end{aligned}$$

Formula (3) shows that the t value of various coefficients and equation overall F value are greater than the critical value at 5% level, the single regression coefficient and overall regression equation in the model

are both significant. But DW value is only 0.0712. It is necessary to test if LRHAPCNI and LNNERL have long-term co-integration relationship, that is, to have the residual term stability test. If successful, there should be a long-term stable equilibrium relationship between the two series.

The second step, use ADF test to have residuals unit root test, results are shown in Table2

**Table 3.** Residuals ( $E_t$ ) stability (ADF) test results

Series	ADF test			
	Test type (C, T, L)	ADF statistics	Threshold	Conclusion
$\Delta E_t$	(0,0,0)	-3.714	-4.8280	Stable

Note:  $\Delta$  represents first order difference; In (C, T, L), C, T, L, respectively represent intercept term, trend term and lag order; ADF test lag order is determined by AIC rule

Table3 shows, the ADF test statistics of  $E_t$  is less than the threshold of significance level 1%, rejecting the null hypothesis, indicating that at 1% significance level there is no unit root, and the residuals are stable, so between variables LRHAPCNI and LNNERL exists long-term stable equilibrium relationship, namely

co-integration relationship.

**Secondly, Co-integration test between variables LRHAPCNI and LRHPCWI**

The same as above, the first step, use OLS to estimate the co-integration equation of the 2 variables:

$$\begin{aligned} \text{LRHAPCNI} &= 3.2188 + 0.6302\text{LRHPCWI} & (4) \\ & (46.5121) \quad (48.5821) \\ R^2 &= 0.98 \quad \text{DW} = 0.5467 \quad F = 2360.223 \end{aligned}$$

Formula (4) shows that that the t value of various coefficients and equation overall F value are greater than the critical value at 1% level, the single regression coefficient and overall regression equation in the model are both significant. But DW value is only 0.5467. It is necessary to test if LRHAPCNI and LRHPCWI have

long-term co-integration relationship, that is, to have the residual term stability test. If successful, there should be a long-term stable equilibrium relationship between the two series.

The second step, use the ADF test to have residuals unit root test, results are shown in Table 4:

**Table 4.** Residuals ( $E_t$ ) Stability (ADF) Test Results

Series	ADF test			
	Test type (C, T, L)	ADF statistics	Threshold	Conclusion
$E_t$	(0,0,0)	-2.7175	-2.6649	Stable steady

Note:  $\Delta$  represents the first order difference; In (C, T, L), C, T, L, respectively represent intercept term, trend term and the lag order; ADF test lag order is determined by AIC rule.

Table 4 shows, the ADF test statistics of  $E_t$  is less than the threshold of significance level 1%, rejecting the null hypothesis, indicating that at 1% significance level there is no unit root, and the residuals are stable, so between variables LRHAPCNI and LNNERL exists long-term stable equilibrium relationship, namely co-integration relationship.

Based on the above analysis, long-term co-integration relationship exists between variables LRHAPCNI, LNNERL and LRHPCWI. Therefore, long-term

co-integration relationship exists in the equation established by the 3 variables.

**(3) Error correction model (ECM)**

The above regression analysis shows that long-term stable equilibrium relationship exists between RHAPCNI, NNERL and RHPCWI, but this relationship is subject to certain short-term imbalance. In order to solve this imbalance, it is necessary to establish ECM containing error correction term<sup>[9]</sup>.

According to the theory of error correction, it is

proper to establish the appropriate error correction model:

$$\Delta LRHAPCNI_t = \beta_0 - \lambda ECM_{t-1} + \beta_1 \Delta LNNERL_t + \beta_2 \Delta LRHPCWI_t + \varepsilon_t \tag{5}$$

In Formula (5),  $ECM_{t-1}$  indicates the error correction term,  $\lambda$  indicates error correction coefficient,  $\beta_1$  and  $\beta_2$  represent respectively the independent variable  $\Delta LNNERL$ ,  $\Delta LRHPCWI$  coefficients,  $\varepsilon_t$  indicates

random disturbance term. This distribution lag model introducing balance error to the first order difference uses OLS to have regression:

$$\Delta LRHAPCNI_t = 0.161 - 0.332ECM_{t-1} + 0.357\Delta LNNERL_t + 0.366\Delta LRHPCWI_t \tag{6}$$

(1.8188)                      (2.9613)                      (2.8938)

$$R^2 = 0.5199 \text{ D.W.} = 2.0178 \text{ F} = 7.9401$$

Formula (6) shows in terms of the short-term fluctuation rule of LRHAPCNI impacted by LNNERL and LRHPCWI, the fitted value of the model is OK, the regression coefficient t value meets  $ECM_{t-1}$  10% level significance, and the other two variables are 1% level significance. D.W. value is equal to 2.02, indicating absence of self-correlation. Error correction coefficient is negative, in line with the reverse correction mechanism. The model suggests that in the long run, some long-term stable equilibrium relationship exists between LNNERL, RHPCWI and RHAPCNI. When short-term fluctuation deviates from the long-term stable equilibrium,

equilibrium state will be recovered at the speed 0.332.

(4) Granger Causality Test

The above analysis shows long-term stable equilibrium relationship exists between the per capita net income of rural households and the number of non-agricultural employed rural labors and rural households per capita wage income, but does this balance have any causal relationship? It is to be verified yet. This section intends to take Granger causality test to test whether there is any causal relationship between these variables. The results are as follows:

**Table 5.** Granger Causality Test Results

The null hypothesis	Lag period	F- statistics	Prob.	Conclusion
LNNERL is not LRHAPCNI Granger causality	2	3.5614	0.0466	Refused
LRHAPCNI is not LNNERL Granger causality	2	1.295	0.2955	Accepted
LRHPCWI is not LRHAPCNI Granger causality	4	2.8105	0.0635	Refused
LRHAPCNI is not LRHPCWI Granger causality	4	5.1459	0.0082	Refused

Table 5 shows in case of Lag Period 2, LNNERL is LRHAPCNI Granger causality, LRHAPCNI is not LNNERL Granger causality, which shows the impact of non-agricultural employed rural labors on per capita net income of rural households is larger, but in turn the impact of the per capita income of rural households on employment of rural labors is not significant. In case of Lag Period 4, LRHAPCNI and LRHPCWI witness two-way Granger causality, showing the per capita net income of rural households and rural household per capita wage income have two-way effect relationship.

**4 Summary and Implications**

Based on related data of China Statistical Yearbook from 1985 to 2012, adopting the co-integration theory, this paper makes an empirical analysis of the long-term balance and equilibrium relationship between the per capita net income of rural households, the number of non-agricultural employed rural labors and rural household per capita wage income and their mutual influences. The results show:

The long-term equilibrium relationship exists between

rural household per capita net income, non-agricultural employed rural labors and per capita wage income of rural households, but witnessing the short term fluctuations which are adjusted to long-run equilibrium. The relationship between per capita rural household income and non-agricultural employed rural labors is: The impact of non-agricultural employed rural labors on rural household per capita income is significant while the impact of rural household per capita income on non-agricultural employed rural labors is not obvious. A two-way Granger causality exists between per capita household income of rural labors and rural household per capita wage income, and the impact at 10% level is significant.

It can give us the following policy implications: First, it is necessary to accelerate non-agricultural employment of rural labors, and improve the income of rural households. Constant and rapid income increase of rural households is bound to have an important positive impact on income of rural households, which requires further elimination of institutional barriers to non-agricultural employment of rural labors, facilitating non-agricultural employment of rural labors. Second, it is necessary to enhance wage income proportion in rural household income, increase investment in human capital in rural

areas, reduce non-agricultural employment cost, guarantee wages of migrant workers and improve the income of rural labors comprehensively.

## Summary

The role of non-agricultural activities of rural labors in their household income is a noteworthy topic. This paper focuses on the long-term and stable equilibrium relationship between the two. Based on 1985 - 2012 Statistical Yearbook data, this study conducted corresponding calculation reasoning and co-integration test of the non-agricultural activities of rural labors and their household income, with the results showing: Long-term equilibrium relationship exists between the two, witnessing short-term fluctuations and a trend of short-term fluctuations to long-run equilibrium.

## References

1. Sun Sanbai, Huang Wei, Hong Junjie: Why free labor migration is so important? - Based on the perspective of intergenerational income mobility [J] *Economic Research*, 2012 (5): 147 - 159
2. Li Shi, Chinese rural labor migration and income growth and distribution [J] *Chinese Social Sciences*, 1999 (2): 16 - 33
3. Cai Fang and Wang Meiyuan, Why does labor migration fail to reduce urban and rural inequality [J] *Economic Perspectives*, 2009 (8): 4 - 10
4. Yan Haokun and Xu Minghui. Rural labor migration and regional economic inequality [J]. *Agricultural Economy*, 2008 (6)
5. Zhu Changcun and Ma Jingzhi, Generalized spillover of urban rural human capital and economic inequality [J]. *China Rural Survey*, 2009 (4)
6. Liao Xianlang, Research on rural labors migration and urban-rural income inequality [J] *population and economy*, 2012 (6): 16 - 52
7. Cai Wu and Cheng Xiaojun, Urban and rural labor migration, urban employment and income disparities - Theoretical and Empirical Study [J]. *Economics and Management*, 2012 (12)
8. Chen Shuwei and Zhang Heping. Empirical Analysis of the Relationship between Urban and Rural Residents Consumption, Government Consumption and Economic Growth in Qinghai Province, 2013 (6): 137 - 133
9. Xu Hangzhou and Guo Yuyan, Co-integration Analysis of Rural Non-agricultural Income and Rural Land Circulation [J]. *Population of China. Resources and Environment*, 2011 (6): 61-66