Journal of Applied Sciences

ISSN 1812-5654
Empirical Analysis Based on Provincial Panel Data Model in Revealing the Role of Urbanization on Urban Employment Rate

Tan Lan
Zhejiang University City College, Hangzhou, People Republic of China

Abstract: Urbanization has influencing effects both on labor supply and labor demand. In this study, an empirical analysis is made to reveal the role of urbanization on urban employment rate. The empirical analysis is based on the data from 29 provinces and during the years between 1995 and 2010. An empirical model is firstly established. And then stability test and Co-Integration test of panel data are carried out. Parameters in the panel date models are estimated. The empirical analysis results show that there has been long-term equilibrium relationship between urbanization rate and employment rate. Overall, urbanization has a negative and significant effect on employment rate. The degree of influencing effect is even large for the medium and western regions than for the eastern regions. The reason lies in that the effect of urbanization on labor supply has exceed the effect on labor demand. To better promote labor demand during urbanization, the most necessary thing for the government to do is to reform the household registration system.

Key words: Provincial panel data model, urbanization, stability test, co-integration test

INTRODUCTION

Urbanization is a historical process of population transition from the rural population into urban population, as well as process of economic structural transition from traditional agriculture society into an industrial society. Seen as in Fig. 1, there has been a steady upward trend in the development of urbanization since 1978. According to statistics, urbanization rate was only 17.92% in 1978, up to 2012, it has soared to be 52.57%, the average annual growth rate is 3.23%. It's also estimated that there will still be a rapid development for urbanization in China in the future and the estimated average annual growth rate is to be 1% point (Jian and Huang, 2010).

Despite of the rapid growth of urbanization in China, China's urbanization level is still lagging behind the world average when considering the level of industrialization. When compared with the world, China's industrialization level is 20% higher than world average, while the urbanization level is almost the same as the world average. This reflects serious lag of the development of urbanization. More specifically, the national income is almost the same as China for Malaysia and the Philippines, but their urbanization levels are above 60%, higher than China. And in Japan and Korea, the urbanization levels are both higher than 70%.

The dual economy is a basic feature of China's economic structure. To achieve the goal of industrialization and modernization, the basic task is to transfer a large number of surplus agricultural populations into non-agricultural populations and promote urbanization. But the transfer process is not smooth until the migrant agricultural labor force can find enough job opportunity in cities. So counties, especially the developing countries are concerned about how to coordinate the positive interaction between urbanization and employment promotion. Researchers have reached different conclusions about the relationship between urbanization and urban employment rate. Some researchers hold that during the process of urbanization, a great number of rural labor force will influx into cities to earn higher wage...
(Lewis, 2003), or to find new jobs (Ranis and Fei, 1961). The huge shift of labor from the countryside to cities will inevitably increase the labor supply and further enhance the pressure on urban employment. So some scholars suggest that the scale and the speed of rural-urban labor force migration should be restricted, so as to solve the problem of urban unemployment (Todaro, 1969).

Some researchers insist that urbanization can create many job opportunities and can expand the demand for labor. Firstly, urbanization can promote aggregate economy. New economic geography theory emphasizes that spatial aggregation of economic activities can bring forward increasing returns to scale and positive feedback effects and thereby promote employment growth (Brakman and Garretsen, 2003; Fujita and Krugman, 2004; Xie et al., 2012). Secondly, urbanization can improve the industrial and the employment structure. Both the development of the service sector and the improvement of the quality of employment can enhance the demand for labor (Yang, 2010; Guo, 2006).

And the other researches claims that since urbanization can enhance both labor supply and labor demand, its total effect on employment rate will be uncertain. If the former effect is dominant, urbanization should have negative effect on employment rate (Mu and Song, 2008) and conversely the effect should be positive (Liu, 2001; Zeng et al., 2007). Fields (1974) holds the view that the effect of urbanization on labor supply is much lower than expected. Firstly, potential rural migrants do not need to go to cities in search of work personally, instead they can find jobs by other ways. Secondly, rural migrants often find jobs in informal sectors, their employment rate are often underestimated. Thirdly, more educated workers will have larger chances to be employed. So appropriate policies can be adopted to lower the negative effect of urbanization on employment (Fields, 1974). In terms of China, the most important influencing policy may be household registration system. The household registration system should be reformed so as to help migrant rural labor force to find new jobs in cities (Cai and Wang, 2010).

In sum, there have been a lot of researches on the relationship between urbanization and employment. Researchers have come to different conclusions due to different research perspectives and different research methods. To reveal the effect of urbanization on employment rate in China more clearly, an empirical analysis based on provincial panel data model is conducted. When compared with the traditional cross-sectional data or time-series data, there are at least three advantages for the panel data. Firstly, thanks to the increased observations, the accuracy of the estimated amount of sampling can be improved. Secondly, consistent estimators of the parameters can be got by panel date models. Thirdly, more dynamic information can be got.

**MATERIALS AND METHODS**

**Methods:** To analyze the effect of urbanization on employment, models based on enterprises optimal employment principle are set up. According to the Cobb-Douglas production function, the output is depended on input of capital and labor force, just as in Eq 1:

\[ Y = AK^{\alpha}L^{\beta} \]  

(1)

Wherein, \( Y \) represents output, \( K \) represents capital input and \( L \) refers to labor force input.

And then, \( \pi \) is used to represent profit and \( p, r \) and \( w \) refer to price of the product, the price of capital and the wage of labor, respectively. Then, \( \pi \) can be calculated to be as in Eq 2:

\[ \pi = p \times Y - r \times K - w \times L \]  

(2)

To achieve the maximization profit, the partial derivative of \( L \) should be equal to 0. Then Eq 3 can be got:

\[ \frac{\partial Y}{\partial L} = \frac{w}{p} \]  

(3)

Then, get the partial derivative of \( L \) from Eq 1 and substitute the derivative into Eq 3. Equation 4 can be got:

\[ L = \beta \times \frac{Y}{(w/p)} \]  

(4)

According to Eq 4, to achieve the maximum profit, the optimal employment should be positively related to the output and negatively related to the real wage. Take natural logarithm on both sides of Eq 4. Equation 5 can be got:

\[ \ln L = \gamma_0 + \gamma_1 \ln Y + \gamma_2 \ln \left( \frac{w}{p} \right) \]  

(5)

Equation 5 is the basic model of the determination of employment. That is, the employment scale is influenced by the economic scale and by the real wage. \( \gamma_1 \) and \( \gamma_2 \) are the elasticity of employment to output and real wage respectively. For example, if \( Y \) is increased by 1%, then employment will change by \( \gamma_1 \times 1\% \), etc.
Define $E$ as the proportion of urban employment workers to total urban population (H). Then Eq. 6 can be got:

$$\ln E = \gamma_0 + \gamma_1 \ln Y + \gamma_2 \ln \left(\frac{W}{P}\right) + \gamma_3 \ln H$$  \hspace{1cm} (6)

When taking into account the effect of urbanization on employment, the following panel data model can be established, shown as in Eq. 7:

$$\ln E_t - \delta_0 + \delta_1 \ln Y_t + \delta_2 \ln \left(\frac{W_t}{P_t}\right) + \delta_3 \ln H_t + \epsilon_t$$  \hspace{1cm} (7)

Wherein, $U_t$ represents urbanization level. The subscript of $t$ represents the corresponding year. The subscript of $i$ represents the corresponding province. Here $\epsilon_t$ refers to the stochastic error for the given year of $t$ and the given province of $i$. $\delta_i$ is the elasticity of employment rate to output. The $\delta_i$ is the elasticity of employment rate to real wage. The $\delta_i$ is the elasticity of employment rate to population. And $\delta_i$ is the semi-elasticity of employment rate to urbanization rate. That is, when urbanization rate is 1% higher, employment rate will change by $\delta_i \times 1\%$.

Based on economic theories, it can be expected that when output is higher, the employment rate should also be higher. So the expected sign of $\delta_1$ should be positive. While the expected sign of $\delta_2$ is not sure. From the perspective of labor demand, higher wage means less profit and less labor demand. But from the perspective of labor supply, higher wage means higher income and more incentive to work. So sign of $\delta_2$ should be depended on which force is dominant. It also can be expected that sign of $\delta_3$ should be negative. When other factors, including economic scale are controlled, the labor demand should be determined, thus higher population means lower employment rate. The sigh of $\delta_3$ is the focus of the study. Theoretical analysis has not give an explicit forecast on the sigh of $\delta_3$. If urbanization have dominant effect on economic development and so force on labor demand, the sigh of $\delta_3$ may be positive. But if effect of urbanization on labor demand is less than the effect on labor supply, then the sigh of $\delta_3$ may be negative. The empirical analysis is aimed to reveal which force is dominant.

Materials: The data comes from China Statistical Yearbook over the years from 1995 to 2010 and covers 29 provinces. Due to missing data, provinces including Xizang and Chongqing are not included. In the selection of indicators, real GDP, that is GDP divided by price level, is used on behalf of $Y$. Average monetary wages of workers also are divided by price level and the ratio is used to delegate real wage, i.e., $w/p$. Urbanization rate and population can be found directly from China Statistical Yearbook. In the data analysis, the natural logarithm of the above data, excluding urbanization rate is calculated. There are two reasons. Firstly the logarithm of the variables does not change the Co-Integration relationship between them. Secondly, Logarithmic variables can eliminate the heteroseasticity which may exist in the raw data.

It is worth noting that due to the complexity of urbanization process, there is no uniform for the index of urbanization. Some scholars use single indicator methods to measure urbanization rate, for example, some use the ratio of urban population to total population and some use the ratio of non-agricultural population to total population. The former indicator is based on the principle of residence. By this principle, whether or not having the urban household registration, as long as the person lives in cities, he is counted as urban population. By contrast, the latter indicator is based on the principle of household registration. By this principle, the person, who can be counted as urban population, must have urban household registration (Ruo, 2006).

By contrast, some scholars insist that single indicator is too simple to describe the urbanization level, so comprehensive index should be used (Zhu, 2012). But which indexes should be chosen in the comprehensive index? So far there is no unified view on this issue. In China Statistical Yearbook, single indicator method is adopted. The urbanization rate is calculated as the ratio of urban population to total population. Although the accuracy of this method is easily affected by statistical coverage of the urban population, however, this method has obvious advantage compared to other methods. That is, the calculation method is relatively simple to be carried out. So far it has become a most widely accepted method. So the indicators of urbanization from the Statistical Yearbook are adopted directly.

Data analysis:
Stability test: According to classic econometric economics, the stability analysis should be done before regression for the panel data model. Non-stationary economic time series often exhibit a common trend, while these sequences do not necessarily have a direct correlation. In this case, despite of the high $R^2$ got by the regression of these data, the result is not of any practical significance. This situation is called spurious regression. Only when the remainder of the data sequence is with zero mean and the same variance after excluding the constant intercept and time trends, can the data be called stable. So in order to avoid the spurious regression
and to ensure the validity of the estimation result, the stability tests on the panel data are carried out. Wherein, in the stability analysis, LLC (Levin, Lin and Chu t), IPS (Im, Pesaran and Shin W-sta), Breitung, ADF (Augmented Dickey-Fuller)-Fisher and PP (Phillips-Perron)-Fisher test are adopted. The stability test results are described as in Table 1.

In Table 1, D(variable) means the first order difference for the variable in the brackets. In stationary test for log(e), D(log(e)), Log(Y), D(log(Y)), Log(w/p), D(Log(w/p)), Log(H), D(log(H)), corresponding test modes are with individual intercept and trends; while in stationary test for Ur and D(Ur), Corresponding test modes are with individual intercept and no trends. The Selection Criteria for the test mode is based on the timing diagram for the corresponding data series. The cross-section numbers in the above tests are all 29.

Seen as in Table 1, most of the test results for the raw data are insignificant. While almost all of the test results for the first order difference are significant at 1% significant level. For example, according to the five different method, the p values (listed in the last column) of Log(e) are 0.0187, 0.2450, 0.8283, 0.5960 and 0.0003, respectively. When p value is larger than 0.1, it can be inferred that the variable is insignificant. Since most of the stability test results for Log(e) are insignificant, it can be concluded that the variable sequence of log(e) is unstable. But when the stability tests are carried out for the first order difference of log(e), i.e., D(log(e)), the p-values are calculated to be 0.0000, no matter which test methods are adopted. It means that the variable sequence of D(log(e)) is stable. The analysis is similar for other variables.

In short, the raw data are all unstable, while the first order difference of all the variable data are stable. So it can be inferred that log(e), log(Y), log(w/p), log(H) and Ur are integrated of order 1.

Co-Integration test: The above stability tests results show that all of the five variables are integrated of order one but it is not certain whether there are long-run equilibrium relationships between the variables. If the long-run equilibrium relationship does not hold, consistent estimated coefficient for the regression equation can not be got. So the next step is to carry out Co-Integration test. In this study, Engle-Granger two-step Co-Integration test method is applied. The first step is to carry out regression and get the residual series and the second step is to do the stability tests for the residual series. If residual series is stationary, then it can be inferred that there is Co-Integration relationship between the variables. The Engle-Granger test results are summarized as in Table 2.

The residual series are obtained by regression of Eq. 7 described in section 2. And the residual series are the estimated value of the stochastic error $\epsilon_r$.

Seen from Table 2, by either test method, the p value is estimated to be less than 0.05. It means that the null
hypothesis is rejected by either test methods at 5% significant level. So there is no unit root for the residual series. And the residual series are stable. Then the conclusion can be reached that, there exists Co-Integration relationship, i.e., long-run equilibrium relationship between the variables in Eq. 7.

RESULTS AND DISCUSSION

Panel data models include three specific forms: Fixed effect regression model, random effect regression model and pooled estimation model. The specific form of panel data model should firstly be determined before estimation of regression coefficients. The selection process can be divided into the following two steps.

The first step is to carry out F test to determine whether pooled estimation model is appropriate. F test results show that the F statistic value is 14.216 and the corresponding p value is 0.00. Therefore, the null hypothesis of pooled estimation model is rejected.

The second step is to carry out Hausman test to determine whether fixed effect regression model or the random effect regression model is appropriate. The Hausman test results show that the Chi-Sq statistic value is 68.37, corresponding p value is 0.00. The test results reject the null hypothesis of random effects regression model, that is, individual effects exist in the model, so fixed effect regression model should be adopted. The regression equation and regression results are shown in the Table 3.

The explained variable is log(e). The explaining variables are listed in Column 1. Regression results for the national data, for the eastern provinces, for the central provinces and for the western provinces are listed in Column 2 to 5, respectively.

Regression for national panel data model: Seen from column 2 in Table 3, the regression results for the nationwide panel data are satisfactory. Adjusted $R^2$ is 0.683, the model fits well, it indicates that the variation of the four explanatory variables can explain 68% of the change in employment rate. F-statistic is large, illustrating that explanatory variables are significant as a whole. In addition, all variables have passed the significance test, confirming the rationality of the selection of variables.

The sign of the coefficients for Ur is negative, suggesting that urbanization has a negative effect on employment rate. Combining the theoretical analysis which has been done in section one, it can be inferred that urbanization has had a big effect on labor supply than on labor demand. It is unfavorable for China’s economy. On one hand, urbanization promotes the transfer of surplus rural labor into cities, increase the pressure on urban employment. On the other hand, the promoting effect of urbanization on economic growth has not been fully realized. The combined result is to be the increased unemployment rate and the intensification of social contradictions. Coefficient for Ur is estimated to be -0.025. It denotes that when urbanization rate is 1% higher, the employment rate will drop by 2.5%.

For the controlled variables, the sign of the coefficient for Log(w/p) is negative, implying that real wage has a larger effect on labor demand than on labor supply. When real wage increases, enterprises have to reduce the employment of workers so as to control the production costs. This effect is larger than the incentive effect of higher wage on labor supply. The sign of the coefficient for Log(Y) is positive, just the same as estimated. Higher GDP means higher output and higher labor demand, so the employment rate will also be higher. But the sign of the coefficient for Log(H) is also positive, opposite to the sign expected. It is rather surprising. After controlling the labor demand, i.e., log (Y), the expansion of the size of the population should affect only the labor supply. Then why the increase in labor supply can raise the employment rate? The possible reason may lie in the age structure of the population. When China has gradually entered the aging society, the expansion of the
Table 4: Production in the three industries as percentage of GDP in 2010

<table>
<thead>
<tr>
<th>Industries</th>
<th>Eastern provinces (%)</th>
<th>Central provinces (%)</th>
<th>Western provinces (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary industry</td>
<td>7.75</td>
<td>12.44</td>
<td>13.37</td>
</tr>
<tr>
<td>Secondary industry</td>
<td>46.56</td>
<td>52.14</td>
<td>48.97</td>
</tr>
<tr>
<td>Tertiary Industry</td>
<td>45.71</td>
<td>35.41</td>
<td>37.66</td>
</tr>
</tbody>
</table>

Further data can be used to explain why such regional difference exists. Seen as in Table 4, in 2010, the average proportions of primary industry production to GDP for the eastern, medium and western provinces are 7.75, 12.44 and 13.37%, respectively. The average proportions of secondary industry production to GDP for the three regions are 46.56, 52.14 and 48.97%. The average proportions of tertiary industry production to GDP for the three regions are 45.71, 35.41 and 37.66%. As is well known that employment absorption capability is the largest for the tertiary industry and smallest for the primary industry. So when compared with the other two regions, the eastern region shows the largest tertiary industry scale, thus its positive effect of urbanization on employment is also the largest. And then, after combining the negative effect of labor supply on employment rate, its negative effect of urbanization on employment should be the smallest in the three regions.

**CONCLUSION**

The main conclusions of the study are summarized as following. There has been long-term equilibrium relationship between urbanization rate and employment rate. Higher urbanization rate leads to lower employment rate. The reason lies in that the effect of urbanization on labor supply has exceed the effect on labor demand. And all of the influencing effects are significant not only for the nation but also for different regions.

For different regions, the degree of effect is different. For eastern regions, the negative effect of urbanization on employment rate is the smallest. The reason for this result is that the tertiary industry has developed the fastest for the eastern regions. This provided the necessary conditions for the expansion of demand for labor.

So in order to improve the employment promotion effect of urbanization, the effect of urbanization on tertiary industry development must be enhanced. In 2010, the urbanization rate in China was 51.27%, slightly higher than the world average of 50.9%. However the development of the tertiary industry is lagged behind the development of urbanization. The tertiary industry of China only accounted for 43.1% of the GDP in 2010, far behind the average level of the world 58.17%.

Then why urbanization has not played the role of promoting tertiary industry development? Studies have shown that the reason may lie in China's urbanization mode. The main difference of urbanization mode between the China and the foreign countries is the household registration system. In China, surplus rural labor can go into cities to find new jobs, but they are not registered residence. Their access to health insurance and children’s
education are restricted. So it is very expensive for these migrant labors to settle down in cities. Therefore, they always work and earn money in urban area when they are young and strong; go back to rural area and consume when they are old or sick. When taking into account the unsustainability of working in cities, these migrant labors are usually much cut so as to save money. Then a strange paradox arises. Urbanization attracts a large number of migrant workers to work in the cities and they create supply of products. But urbanization has not encourage these migrant workers to consume in cities, so it restricted demand for consumption and for service in urban area. As a result, domestic demand is inadequate and the development of tertiary industry in cities is also restricted. It in turn restricts the growth in the job opportunities.

Therefore, in order to better promote employment growth during the urbanization process, the government should firstly reform the household registration system. Migrant surplus rural labor should not only work in cities, but also live in cities. Then they will create both supply of and demand for production. And this is important for the development of tertiary industry in cities and for the increase of labor demand.

ACKNOWLEDGMENT

This study was supported by the Construct Program of the Key Laboratory in Hangzhou, by Finance Leading Academic Discipline in Zhejiang University City College and by Hangzhou Regional Economic Research Base. Financial support from the MOE Project of Research of Humanities and Social Sciences at Universities (Project No.: 10YJC790231) and the Zhejiang Provincial Natural Science Foundation of China (Project No.: Y6110353) is gratefully acknowledged.

REFERENCES


