

Reply to referee 2

Thanks a lot for the comments and suggestions from the referee. We revise the paper by following the referee's report strictly.

Q1: It would however be worth to explore another factor analysis method

Wilson and Cooper (2010) explain other inconveniences of this method and advise using Velicer's minimum average partial (MAP) method that, besides producing a one factor solution to a dataset, it also calculates an associated index based on the average squared residual correlations of that one-factor solution. This method has many advantages (see Wilson and Cooper, 2010) that would be worth to explore and use in this kind of studies. Free factors analysis programs, such as FACTOR (Lorenzo-Seva and Ferrando, 2006) could be used for this method.

A: We are so appreciated for this suggestion to use MAP method. But as the referee said, PCA is still the most popular method for this kind of studies. Furthermore, the factor analysis isn't our core interest in this paper. So, could we keep the PCA results here? We may use the MAP method in my next study using this data.

Q2: In the factors description the authors find an ambiguous relationship between integration and duration of stay (although the overall integration level is positively related to duration, they still find that migrant workers with shorter duration enjoy higher integration in term of social interaction, psychological integration and sustainable development). Could this result be interpreted as a fatigue (or crisis) of integration that overcome the willingness to integrate if the time is not sufficient yet to acquire all the tools for a proper and more stable form of integration? Which of the five elements of social integration we speak about for those subjects that integrate in the short run and for those that integrate in long run? Is there any difference in the composition of their integration?

A: The longer a migrant worker stays in city, the less he/she integrated. This result could be interpreted as a fatigue (or crisis) of integration that overcome the willingness to integrate as duration grows. The possible explanation is that migrant worker with longer duration are more likely to isolate themselves from local society because they are more realistic. They gradually find it so difficult to survive in the city and city life isn't so attractive anymore. For those that integrate in the short run, social interaction and motivation of settlement play a more important role on their overall integration. For those that integrate in the long run, social distance and cultural integration are more significant for their overall integration because they are more urbanized and more difficult to return.

Q3: The "marriage" control variable means marriage to the local person? Three forms of marriages (or even partnerships) could have different effects: 1) to a local to the city, 2) to a foreigner (to the city and to the observed subject), and 3) to a person from their home region. It is slightly different from the social contacts variable as being with someone does not directly mean having wider social life, although it has potential to have even stronger effect. It would also be interesting to control for the number of family members that live or have ever lived in a city (Chinese city or foreign - to see to what extent the family has experience to overcome the cultural barriers) that they have contact with.

Is not there any collinearity (or causal relationship) between the control variables (e.g. income and/or age) and consumption and its components?

A: First, to investigate on different forms of marriages and to control for the number of family members that live or have ever lived in a city are of great interest and importance to make a better understanding to the issue what affect the social integration of migrant workers. What a shame is that we have no such information. But this suggest inspires me to further this survey in the future. Second, there is no severe collinearity concerns, otherwise, STATA will drop the variables who suffer from collinearity. The main source of collinearity may come from income since there should be a high correlation between income and consumption. But the VIF of income is 3.66, which is much smaller than 10, showing that the collinearity of income isn't our major concern. The possible reason is that the high correlation between income and consumption could be mitigated to some extent by transferring income variable to dummy variable.

Third, we must admit that our baseline results suffer from endogeneity of consumption. The referee suggested to use lagged consumption or indicators reflect the consumption of correspondents in previous periods should be good instruments. What a shame is that we just acquire a static data, which means we have no information on correspondents' previous or lagged consumption. However, we construct a different instrument using other 8 correspondents' consumption with nearest consumption level. Some relevant tests prove that this instrument is well-performed. The results of ivregress in column 1 of table 9 is in line with the baseline study.

Q4: The authors suggest an explanation of result: "... in comparison with the effects of other factors, the effects of consumption level are found to be relatively low. This is because most migrant workers need to compress their consumption due to the dilemma of low income and heavy family burden, which leads to low consumption level". Is this an appropriate explanation of the relatively low significance of the effect of consumption on social integration? The authors explain the negative and significant effect of the interaction term consumption*income as follows: "We might attribute this result to the diminishing proportion of consumption over income, causing the decreasing utility of social integration for migrant workers derived from consumption with the increase of income". Is this what the survey data show?

A: We think it is an appropriate explanation of the relatively low significance of the effect of consumption in China. Most migrant workers need to support their family financially, which forces them deposit more and reduce the expenditure. According to a report released by China NBS, the total of remittance migrant workers send home is more than 200 billion Yuan (about \$30 billion) in 2006. For the second question, it is a pity that we only have dummy income rather than exact income, so we couldn't prove it through the survey data. We just say this according to the classical consumption theory.

Q5: The article is written in a clear way when it comes to its specific parts. The structure is less standard however it is not to say severely less „reader friendly“.

Description of the consumption structure dataset creation would be more reasonable not to "hide" in the 3.4. "Specification" part and to describe it clearly in another (former) part of the paper.

A: We adopt this suggestion and put the description in the measure part of the paper in 3.2.2.

Q6: The authors did not explain how they created the consumption behavior data.

A: We calculate the average scores of different types of consumption behavior for each

correspondent, and then judge which type of consumption a correspondent belongs to according to the highest average score. The overall sample is split into 4 subsamples according to the highest score of each correspondent.

Q7: It is also important to revise the English, as the article is full of errors in terms of grammar.

A: The writing of the paper is improved dramatically through proofreading service.

Q8: Idea for a following study

Having five-dimensional definition of social integration at hand, it would be interesting to explore the same effects on the disaggregated data of social integration.

A: We do it in the current paper by providing some robustness checks using disaggregated data of social integration. and the results are found to be consistent with our baseline study.