Referee Report „Occupational Upgrading and the Business Cycle in West Germany"

This paper investigates whether firms respond to economic shocks by adjusting their hiring standards rather than by adjusting the wage. More specifically, the authors analyze whether, within detailed occupations groups, firms hire less skilled workers in upturns and more skilled workers in downturns. The authors also investigate the cyclicality of 'occupational composition wages', by testing whether workers are more likely to upgrade (moving from a low to a high wage occupations) in upturns, and to downgrade in downturns. None of the proposed tests are new; this paper replicates the analysis by Devereux (2002) for German social security data. Devereux's paper was published in the journal 'Labour'. This begs the question: How far can a paper that replicates a paper that was published in 'Labour' go? Personally, I think that Devereux's paper is under-placed. Moreover, I find the comparison between the US and German labour market very interesting, and this is a contribution in itself. If I had to referee the paper at, say, the same journal as the one where Devereux's paper came out, I would propose a revision—although a major one.

Also: one of the authors (Johannes Ludsteck) has a paper that seems very related to the current paper, but is not cited. This is unacceptable. The current paper needs to highlight what it adds relative to the older paper.

Part of the writing is quite awkward, and the structure of the paper could be much improved. But the empirical analysis itself is competently and carefully done.

I have the following suggestions to the authors, meant to improve the paper:

1) The structure of the paper is a bit messy, and is sometimes difficult to follow. In order to focus the paper, why not first describe the results (regarding the hiring standard and the occupational composition wage) for Germany, and then compare them with the US findings. Here, it would be useful to include the findings for both countries into the same table. Finally, there could be a discussion section that proposes several explanations why the results for Germany differ in magnitude from the US findings, and rules out explanations that do not work (such as collective bargaining). For the moment, the paper first discusses results, then explanations, then results again, and so on, making it difficult for the reader to follow the main arguments.

2) One proposed explanation for the lower responsiveness of hiring standards in Germany is the German apprenticeship system, which may lead to entry barriers into occupations for workers who did not complete an apprenticeship in that occupation. I like this explanation. However, I find the proposed test for this explanation (restricting the sample to apprenticeship occupations) very weak. I am not even sure the authors should present this test.

3) My biggest concern regarding the empirical analysis is the noise in the education variable. As the authors point out, it may well be the case that firms classify a worker as low skilled if he is currently performing tasks that a low-skilled typically performs, although he is in fact a skilled worker. This in itself could explain the lower magnitudes in the German data. I am also concerned about simply dropping workers with missing education. This could produce a severe selection bias if the education variable is predominantly missing for less-skilled workers. The education variable could be much improved by using the longitudinal aspect of the data, and I strongly suggest the authors to repeat their analysis for a 'cleaned-up' education variable.

4) The data section could be much shortened. It contains quite a few not too important details that could be moved to an appendix or dropped entirely. (For instance, it is not

necessary to stress that firms are not sanctioned if they misreport certain characteristics, but that the quality of the data (with the exception of the education variable) is nevertheless good. The data description should highlight that the data, sample and control variables are very similar to that used by Devereux, and point out the remaining differences.

5) Regarding the result section: Since the goal of the paper is to replicate Devereux's analysis for Germany, the authors should use the exact same methods as Devereux whenever possible. I would therefore focus on the linear model in Section 4, and considerably shorten the discussion about the grouped probit model. Since the grouped probit model has some advantages over the linear model, the authors may also report these findings. But the main point should be that the linear model is a good approximation. The same goes for Section 4.

6) I found the introduction quite difficult to read. It could be considerably shortened, focusing on what this paper actually does: a comparison of the responsiveness of hiring standards and occupational composition wages to the business cycle between the US and Germany, and what we learn from this analysis. The general motivation in the first two paragraphs of the introduction seems out of place.