

Referee Report on MS 1420 entitled

“Modeling Bank Asset Quality and Profitability - An Empirical Assessment”

This paper applies a panel data consisting of five groups of banks in Indian banking industry from 1997-2009 to discover the determinants of bank asset quality and bank profitability. It contributes to the literature on Non-performing assets and bank profitability in emerging economies with some interesting results. The empirical specification is reasonable and the basic results are sound.

However, the paper still suffers from several weaknesses that need to be fixed.

- The structure of this paper seems in some mess and needs to be better organized. For instance, Part III, Part IV, and Banking Profitability in Part V could be combined, and serves as literature review. Part II and Part V (Asset Quality in Indian Banking) could be put together, and serves as background. There are also some typos through the paper. For instance, the first sentence in Part V is weird.
- Moreover, it is better for the paper to focus on Indian banking industry and stress that it contributes to literature on emerging economies. Currently, the abstract and the introduction seem like that the analysis is a panel of emerging economies, causing some confusion, and does not read fluently as well.
- In the regression estimation for analyzing the determinants of asset quality in in table 5A and 5B, the dependent variable is gross non-performing assets (GNPA), which is very likely to have something to do with the total assets of a bank (ASSET). However, the size of the bank is not controlled in table 5A or 5B. To fix this problem, the authors could either use the ratio of NPAs to ASSET as the proxy for bank asset quality, or control ASSET. Considering some control variables, the ratio may be a better indicator.
- In the results report, it is not clear how large the sample size is, the frequency of the data, and how many banks in each bank group. As a result, the referee is even not sure about whether each observation actually represents a bank or a group of banks in the regressions.
- In analyzing the determinants of asset quality, it needs to provide some explanations for adopting different explanatory variables in table 5A and 5B. GDPGR is indeed an alternate variable of IIPGR. But INFLA should also be controlled in table 5A. Similarly, CDR shows up in 5A, but not in 5B, while LR, MCAP shows up in 5B, but in 5A. In addition, it is useful to provide some explanations as well for the differences between the determinants of bank asset quality and bank profitability.

- Some results and the analysis in table 5-7 are not reasonable. For instance, in table 5A, the sign of credit-deposit-ratio (CDR) is opposite to expectations, as CDR indicates the aggressive lending activity of the bank and is predicted to have a positive effect on the NPA levels. It is also hard to explain the significant and positive signs of OER and ROI, as well as the insignificance of GDPGR in all tables. In addition, unlike the authors' claim, the findings about rural branches are consistent with, rather than contrary to, the general perceptions as the regression results of RUSUBRA imply that it will positively affect the NPA levels.
- It is better to extend the period of 1997-2009 to 1991-2014 if possible. As shown in Figure 2, there seems to be a time trend in NPAs during the period of 1997-2009. Moreover, the authors could benefit a lot from the analysis of a full period of 1991-2014 and three sub-periods of 1991-1996, 1997-2009, and 2009-2014, respectively. The comparisons between the three sub-periods would be very useful and interesting. People would be particularly interested in the NPAs and bank profitability after the global financial crisis happened in late 2008 (or early 2009). This advice also applies to the tables and figures in Part II.
- The paper uses two estimation methods, namely Panel Least Squares and GMM. For the panel least squares estimation, it seems that the authors did not control fixed effects or year effects. At least, it is better to add a regression that controls all group dummies if it is not appropriate to control fixed effects. For the GMM estimation, the referee is wondering if the method is appropriate here as there are no time lags of dependent variables in the regression equations.
- It is more meaningful to report the results of simultaneous equations as baseline results, not as the robustness tests. One reason is that the endogeneity problem does exist in the paper, as bank asset quality is one determinant of bank profitability. While in the robustness tests, the authors could add in a similar robustness test for the simultaneous equations as to other regressions.

Overall, this paper has potential contributions, but might need a major revision as suggested above.