Comments on "A Replication of 'The long-run impact of foreign aid in 36 African countries: Insights from multivariate time series analysis' (Oxford Bulletin of Statistics and Economics, 2014)"

This is a reasonable replication study. It deals with an issue of some significance, and, according to Google Scholar, the replicated study has attracted a fair number of citations at 103 over a 5-year period. The motivation for working with alternative data sets is stated clearly even though there might be some overstatement in the sentence about "far-reaching policy implications…(of) the results of the literature on foreign aid effectiveness" (p. 2). The literature review is largely appropriate and relevant to the replications done by the author. The comparison of the data sets (pp. 6-14) can be useful to other researchers also.

Based on the criteria (Table 3) for conclusion about consistency with the JMT study, which seem reasonable, the replication results are summarized in Table 4, which shows that while the replications using JMT models and PWT 6 data yield results largely consistent with JMT, estimates from the three alternative data sets lower the proportion of consistent cases to about two-thirds. The author nevertheless states (p. 19) that notwithstanding the relatively large variation in individual coefficients and country-wise inference, "the overall conclusion of aid effectiveness…remains unchallenged throughout all datasets; it is, however, slightly weakened". What is perhaps meant is that in a majority of cases, aid is found to be "effective" in the sense of (significantly) increasing GDP or investment or both.

The additional exercise undertaken by the author is to respecify the models on the basis of the data sets in the spirit of "allowing the data to speak freely". The respecification is done for 12 countries and covers (a) trends and dummies, (b) lag-length and (c) cointegration rank. The author considers these issues thoroughly (pp. 20-31) and also does misspecification tests. The chosen model specifications are summarized in Table 8, and the test outcomes are summarized in Table 9. The exercise indicates that the respecified models indicate less consistency across the data sets than the estimates from the JMT models.

The author finally (p. 36) notes that choice of data sets can have substantial impact on the results from time-series analysis, and this needs to be addressed in future research. Also, alternative estimation procedures should be used to judge the robustness of the results to data and estimation methods. The author suggests that in view of the fair degree of consistency in the results from four different data sets, the pessimistic tone of several scholars in regard to the quality of development data for African (and other) countries is not legitimate.

Given my limited understanding of time-series methodologies, including CVAR models, the replications seem to have been done properly and are useful in indicating (a) sensitivity of the estimates from the same models to different data sets, (b) additional sensitivity of the estimates from respecified models, and (c) the overall conclusion of a positive effect of foreign aid on macroeconomic variables despite the fairly sizable inconsistencies across the data sets and due to the model respecification. My thoughts on the JMT study and the replications are thus of a somewhat incidental and minor nature and are indicated in the following paragraphs.

1. Unless the exercise is directed toward time-series experts, it may be indicated, for the benefit of a lay reader, how the models enable one to judge the impact of foreign aid on economic growth, which is the theme pursued in most of the relevant literature. As far as I can see, the matrix in equation (2) on page 14 refers to the impact of ("shocks to") foreign aid on the level of GDP and not on its rate of growth. If

the growth impact is reflected in some other estimates, those might be mentioned. The effect of foreign aid inflows on the level of GDP appears conceptually straightforward and the usefulness of the estimates might be limited.

2. Inferences in the JMT study and the replications are based on the signs and significance of the estimates, and magnitudes of the estimates are apparently not considered. It seems that magnitude of the estimated effects should be as important as their signs toward an indication of the effect of foreign aid on growth or GDP (or investment). A mere count of the signs and significance of the estimates presents an incomplete picture and is not a reliable basis for an inference on the role of foreign aid in the level or growth of output. In most traditional regression models, signs and magnitude of the estimated parameters are reported and often considered. If time-series procedures, including CVAR, are not able to provide reasonable estimates of the magnitude of effects, that would be a significant drawback of such procedures.

3. An appropriate inference about the merit of foreign assistance should perhaps consider not merely the absolute magnitude of the effect on growth or GDP or investment, but also some measure of the "rate of return" in the spirit of Arndt, Jones and Tarp (WBER 2016). If the effect of aid is positive, but "small" relative to the aid, and the return on aid is low, that would tend to attenuate one's "optimism" about the effect of aid.

4. An inference about the "favorable" effect of foreign aid based on its positive role in growth or GDP (or investment) of the recipient country reflects an incomplete picture in another sense also even when the magnitude of the effect has been considered. Aid flows from donors to the recipient countries are likely to have some "opportunity cost" and the favorable effect on growth, GDP or investment in the recipient country should perhaps be considered along with its opportunity cost for a more complete picture of the consequences of foreign aid.

5. The JMT work and the replications seem to be largely academic exercises without major "policy" implications. There is apparently a consensus in the international community that aid helps developing countries, particularly the least-developed countries which dominate the samples used in the JMT study and the replications. The MDGs and the SDGs seem to bemoan the fact that ODA is well below the targets. For example, SDG Report for 2018 (p. 30) notes, "The United Nations has set a benchmark for ODA contributions of at least 0.7 per cent of a country's Gross National Income (GNI). However, only five DCA countries…met this target. Overall, ODA remained at 0.31 per cent of GNI in 2017." Therefore, in view of the international consensus on the need to increase ODA, with the implication that such an increase would help the recipient countries, there is no real "policy" issue.

6. JMT's statement (p. 153) about the "polarized nature of the aid debate", which partly motivates their study, and the expression "notoriously controversial aid-effectiveness debate" in the present paper (p. 1, abstract), seem like exaggerations. My impression from the recent literature on aid is that there is a kind of consensus even in the academic circles about the positive effect of aid on growth. For example, Arndt, Jones and Tarp (2016, p. 446) note, "In recent years, academic studies have been converging towards the view that foreign aid promotes economic growth" and "the large majority of up-to-date empirical studies in the economics literature have found positive impacts. More precisely, the full range of independent studies published since 2008 based on cross-country growth regressions report comparable results for the marginal effect of aid on growth." Therefore, the predominant view in the recent academic literature also appears to be inclined toward a positive effect.

7. The discussion of the model respecification procedure (pp. 20-31) for 12 countries appears too long. While it is useful and even necessary to make the procedure clear and transparent, the discussion, which takes up nearly one-third of the space in the text, may hinder comprehension of the replications. Perhaps the author could retain the core parts in the text and include the rest in an appendix.

8. Foreign aid does not seem like a homogeneous entity. It can take several forms, and each form of aid may have a different effect on the recipient country. The composition of aid may thus be a significant factor in the direction and magnitude of its effect. Working with total aid misses the role of its composition.

9. The macro variables studied by JMT and in the replications (GDP, I, C and G) may be related by a national-accounts identity, especially if the economy is treated as closed, and simultaneous estimation of the coefficients of these components in a VAR might be problematic.

10. If adequate number of observations is available, one major attraction of time-series methodologies is that these enable one to judge the effects in an individual country, which cross-section and panel procedures generally cannot. However, if one's interest is in the overall position for a group of countries, the uncertainty in estimates from time-series procedures seems to be of the same kind as in cross-section and panel models, and can arise from variations in the coverage of countries, data sources, estimation methods, period studied and variable proxies. Moreover, as the discussion on pp. 20-31 indicates, there may be greater uncertainty in the time-series procedures due to the researcher having to make several kinds of choices that do not need to be made in cross-section and panel models. For the present case, the author could perhaps provide a more direct indication of the degree of consistency in the inferences from the four data sets for the JMT models and the respecified versions. That could be done by including a summary table indicating, for each of the 12 countries, the conclusion about the effect (positive, negative or insignificant) on GDP and investment in PWT6, PWT7, PWT8 and WDI data sets in terms of the JMT models and the respecified versions. This would show the number of countries (out of 12) for which estimates from the four data sets for the two sets of models yield the same inference. It would be some modification of Tables 4 and 9 and might be done for the positive prior or for both positive and negative priors. Although variation in the magnitude of the effect might not be available, at least the qualitative conclusions could be compared across the estimates for GDP and investment for each country. Table 9 gives the impression that there might be no country for which all estimates yield the same inference.

11. The notation in Table 1 (p. 11) may be made more clear by indicating what exactly Δ GDP, Δ Investment, Δ Consumption and Δ Government stand for. The text refers to "annual GDP growth rates, and the shares of investment, consumption and government expenditure", and it is not clear whether these are rates of growth of the variables or denote absolute changes or something else.

As a minor point, "Statistics and Economics" in the title may be changed to "Economics and Statistics".

I would like to add that most of the foregoing thoughts are not intended to be critical of the work and might be unimportant in a replicatory study. However, these seem relevant to the substantive issue explored in the JMT paper and the replications and may indicate the significance of the exercise in a broader context.