**Referee Report on "Normalized CES supply-side system approach: How to replicate Klump, McAdam, and Willman (Review of Economics and Statistics, 2007)" submitted to the Special Issue on "The Practice of Replications" for *Economics: the Open Access, Open Assessment E-Journal* (MS 2348).**

This paper sets out a replication plan for the influential paper by Klump, McAdam, and Willman (KMW). According to the Editor's instructions to contributors, the replication plan is required to provide:

(i) a general discussion of principles about how one should do a replication,
(ii) an explanation of why the "candidate" paper was selected for replication,
(iii) a replication plan that applies these principles to the "candidate" article, and
(iv) a discussion of how to interpret the results of the replication (e.g., how does one know when the replication study "replicates" the original study).

Narrowly interpreted, the paper meets these four requirements. Despite that, the overall value-added of such a replication exercise would be relatively limited, as discussed below.

For (i), the authors briefly review some background literature on replication in economics and conclude (p.2) that "the replication study should shed some light on how 'generalizable' the results from the original study are". This suggests that their proposed replication will go beyond a 'pure replication', in Hamermesh's (2007) categorization, or 'verification', in Clemens' (2017) classification. However, in the data section (p.6), it is stated that "the data sample period should be from 1953 to 1998" (the span of data in KMW's study), and the authors comment that "we believe that a verification approach should be used. The data should be remeasured using similar methods to verify and rectify any potential measurement errors or coding errors in the original study". It is not explained how the same vintage of data used by KMW can be recovered, an important point already discussed in detail by Referee 1 for this paper, or whether and how 'remeasurement' would be handled. However, Stewart's (2017) replication of KMW is based on data obtained from Alpo Willman, one of the original authors, so the data should be directly obtainable. The authors then add that the sample period should be extended to include the most recent data. This is a sensible suggestion but, apart from a brief footnote (fn.5), there is no consideration of whether there are any problems obtaining comparable data on all the variables for the extended period. In addition, all the material on the interpretation of results in section 4 relates directly to the original KMW results (pp.9-10 is based on KMW's pp. 188-189, except written from an ex ante perspective). It would be useful to have some explicit discussion of how to assess the results for the extended data set, given the possibility of variation over time in key parameters, i.e., have the parameters of the production technology changed over time and how would this be assessed?

For (ii), the authors include a brief section on why they chose to replicate KMW's paper; this makes a case for the importance of KMW's paper, but it could usefully be extended by saying a little more about KMW's influence and later developments.

For (iii), the replication plan works through the steps in KMW's original analysis, including their model, data and estimation methods. This summarizes KMW's paper, with considerable space devoted to the equations in KMW's system of equations (eqs (3)-(6) on p.5) and the different model specifications fitted by KMW, corresponding to the different columns in KMW's Table 1. Indeed, the authors reproduce KMW's Table 1in full (p.8).

Practically all the discussion in section 3, including section 3.2, is directly paraphrased from KMW.

For (iv), the tone of section 4 is that a successful replication is one that yields the same or very similar results (in terms of size and significance of parameter estimates) to KMW's paper, mimicking all key aspects of the original (see especially p.9). However, at the top of p.8, the authors state that "a successful replication of this study should possess parameters that are mostly significant for all specifications as well as report the log determinant (or log likelihood) for the system and the Augmented Dickey-Fuller (ADF) test for the residuals for each equations". This needs to be rewritten as the first part puts sole emphasis on statistical significance; a set of mostly significant parameter estimates but associated with markedly different parameter values from the original would not constitute a "successful replication". Similarly, merely reporting log determinant or ADF values would not be consistent with a "successful replication" if they led to substantively different conclusions compared to the original.

Overall, the approach proposed is a routine verification exercise with a few comments hinting at extension to an updated dataset. How much can be learned from such an exercise? For the verification exercise on the original data, there is little scope for value added given that Stewart (2017) has recently replicated the original results in KMW, using different software (TSP instead of RATS). Stewart's replication makes the added contributions of nested testing of the models, testing the special case of logarithmic growth of technology, and exploring the properties of the likelihood function in the region of $\sigma = 1$. Where the current replication plan could contribute new insights would be application of the model and estimation methods to an extended data set, but this part of the replication plan is underdeveloped and needs to be discussed in more detail.

**Minor points**

p.1, para 1, line 3: delete 'of' before "a series of …"
p.1, para 2, line 3: insert 'the' before 'ideal direction'
p.1, para 3, first sentence: incomplete sentence
p.1, para 3, line 7: delete 'for' before 'sampling error'
p.1, para 3, lines 8 and 9: "Chang and Li (2017)" instead of "Chang et al. (2017)", and "Anderson and Kichkha (2017)" instead of "Anderson et al. (2017)"; also fix the references on p.11.
p.1, para 3, last line and p.2, line 1: Insert apostrophe in 'authors' (two cases), as in the original.
p.1, fn.2, line 2: word(s) missing? Also, line 5: How does an extension test eliminate the "effect of outliers by using the identical computer code on the new data"?
p.2, section 2, para 2, line 1: Correct typos in 'Consant' and 'Substituition'.
p.4, line 4: In eq. (2), $\pi$ is the capital share of income, not the labor share. (Also, move 'and' from before *K* to before $\pi$.)
p.4, para 2, line 5: Again, $\pi$ is interpreted as the labor share of income; also, p.5, line below eq. (6).
p.4, para 2, line 7: Delete apostrophe in "author's".

p.4, para 2, line 9: As it stands, “and this would be the case” is ambiguous – sample averages coincide with the implied fixed point only if the functional form is Cobb-Douglas with constant growth of technology.
p.5, last para below eq. (6): Delete ‘of’ before “is denoted” and delete ‘the’ before “Box-Cox”.
p.6, section 3.1, line 1: ‘principal’ instead of ‘principle’
p.6, last para of section 3.1: fix typos: ‘similiar’, ‘mearsument’, ‘extendend’
p.7, section 4, lines 3 and 5 (and other instances): ‘log determinant’ instead of ‘log determinate’
p.7, fn.6 (also reference list on p.13): The date for Stewart’s replication of KMW is 2017 not 1938!
p.8, lines 1-4: fix typos: ‘dermeninate’, ‘equations’
pp. 8-9: Note ‘g’ should be included in the table notes, not as a separate footnote (but why ‘g’?).