

Response to Referee 2's Report on Normalized CES supply-side system approach: How to replicate Klump, McAdam, and Willman (2007)?

Thank you for your insightful comments, all of which will be addressed to improve the paper. Responses are given below in blue (following the original comments).

Comment 1:

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?

Response: To address your comment above (which overlaps with Referee 1’s comment about the NIPA vintage data), we will include a discussion that nuances our original message. This will be added in section 3.1 last paragraph. Specifically, we will state:

“When replicating this study, 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 should be noted that it is impossible to recreate the original data set since new vintages of NIPA data differ. While differences across vintages are small, differences for capital stock data may be larger. Thus, according to the verification approach, a replication may yield results that are reasonably close to original but not exact replications. This doesn’t pose so much of an issue. However, there may be difficulties in clarifying whether differences are real or a reflection of some unnoticed mistakes in recreating the data that is used in replication. If the sample period is extended to the most recent data available, one may or may not reproduce results as in KMW. Since, the results in KMW may be time-sensitive.”

#### Comment 2:

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.

Response: We appreciate this comment and agree with you that a couple of later developments should be added to our paper. Section 2, last paragraph specifically will state:

"Leon-Ledesma et al. (2010) use the normalization approach to identify the conditions under which joint identification of capital-labor substitution elasticity and technical biases in production are feasible and robust and also show that this technique is superior to single-equation estimation approaches especially when merged with "normalization". Further, Klump et al. (2012) presents an exhaustive survey assessing the intrinsic links between production (as conceptualized in a production function), factor substitution (as made most explicit in Constant Elasticity of Substitution functions) and normalization (defined by the fixing of baseline values for relevant variables). All these studies justify more extensive use of CES production functions in dynamic macroeconomics."

#### Comment 3:

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.

Response: We appreciate this comment and agree with you that the value of the parameter estimates should also be emphasized. Thus, we have modified the first full sentence on page 8 to read as follows:

"Thus, a successful replication of this study should possess similar parameter values 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 equation".

#### Comment 4:

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  $\beta = 1$ . 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.

Response: While Stewart (2017) recently replicated results in KMW using a different software, the data used for their replication was obtained directly from one of the authors of KMW. In replicating studies such as KMW, data integrity is of the utmost importance. To emphasize the importance of data integrity, the authors themselves take care to note, in their abstract and concluding remarks, that they put "a high emphasis on data consistency" (p. 183). Thus, a successful replication of KMW must remeasure the data used in KMW as minor measurement or coding errors may lead to different results than presented in KMW. It should be noted that it is impossible to recreate the original data set since new vintages of NIPA data differ. While differences across vintages are small, differences for capital stock data may be larger. This issue has been acknowledged earlier, see response to comment 1.

Comment 5:

### 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"? [The computer code cannot take care of outliers even with the original dataset.](#)

p.2, section 2, para 2, line 1: Correct typos in 'Consant' and 'Substitution'.

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, 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 referencelist 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'?).

**Response:** All the above typos are fixed.