

Response to Referee 3's Report for "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:**

This paper proposes a replication plan for Klump, McAdam, and Willman (2007). Klump, McAdam, and Willman (2007) estimates the supply side of the U.S. economy from 1953-1998, using a constant elasticity of substitution (CES) production function. They find that the elasticity of substitution between labor and capital is significantly below unity, and that the growth rates of technological progress show an asymmetrical pattern, which are consistent with prior criticisms of the use of the Cobb-Douglas production function. To replicate this paper, the authors propose a verification approach, "where the data will be remeasured using similar methods to verify and rectify any potential measurement errors or coding errors in the original study." The authors also plan to extend the sample period to incorporate the most recent data available.

My overall evaluation of this replication is that, as is, their verification exercise has the potential to make some scientific contribution to the literature. However, I believe there is potential for a significantly more meaningful contribution if the authors broadened the scope of their replication plan to assess the conditions under which the results are likely to hold, or to include additional robustness tests. Indeed, this tension exists within the authors' own evaluation of the best practices for replication. For example, the authors argue that, "We also believe that replication studies should go a step further in assessing the conditions under which the original study are valid. In other words, the replication should shed some light on how "generalizable" the results from the original study are." They argue that replications should be "purely scientific in nature," which refers to the typology developed in Hammermesh (2007). Hammermesh distinguished "pure replications", which use the same sample, model, and estimation methods, from "scientific replications," which use a different sample, different population, and perhaps similar but not identical model employing alternative theoretical or conceptual approaches.

While the authors argue for a broader scope for replications in their description of best practices, their actual plan is almost exclusively a narrow verification exercise. The authors will be estimating the same model, using the same measures (although re-measured to ensure no coding errors), over mainly the same time period. As such, the only flaws in the original study that could be detected would be equivalent to typos or coding errors. However, as the authors themselves acknowledge, the more significant contribution would be to also assess the conditions under which we would expect the original results to hold.

Response: We agree with the original authors that estimation approach used in their study is effective and widely in this literature. In addition to discussing the steps for a pure

replication which entails estimating the model using the same sample, we do recommend changing the sample period and acknowledge that this new estimation may significantly impact model estimates. On page 6, we state “the sample period should be [extended] to the most recent data available as the result should not be sensitive to a larger sample period.” Since the estimation technique used is sensitive to the baseline values, changing or extending the sample period could have a significant impact on the model estimates. Thus, a replication of the original study with a different sample period would amount to much more than the detection of typos or coding errors, i.e. one form of robustness check.

To further address this concern, 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:**

Unfortunately, due to the limited scope of the exercise, there is no discussion of alternative tests that could occur which would shine light on these more significant issues. The authors begin to approach these concerns when they suggest expanding the sample by including newer data. However, their concerns are narrowly tailored to whether the original results would hold up with the inclusion of the new data. It would be more interesting to hear how the authors think the new data could also be used to verify older results. For example, the newer years could possibly be used to evaluate the previous results by assessing how well the previous results predicted out of sample, or through some other type of structured comparison between older and newer data.

Response: We agree with the referee that there are other interesting robustness checks that can be used.

To further address this concern, we will state the following in our Data section:

“The production function implicitly assumes that elasticity of substitution is time invariant, and the researcher replicating the original study should evaluate the sensitivity of this parameter. In particular, the researcher replicating the original study should also change the sample window to determine if the elasticity of substitution changes substantially with the window size. This can be done with the original data set or with an

extended data set.”

**Comment 3:**

Additionally, as a narrowly tailored verification exercise, there is no discussion of additional robustness tests that could shine further light on the general conditions under which the results would hold. The original paper grew out of dissatisfaction with the use of a Cobb-Douglas production function to describe the U.S. economy. The authors could potentially add value by designing some type of test which could assess some of these earlier dissatisfactions, and confirm whether and under what conditions the Klump, McAdam, and Williams (2007) methodology successfully addressed these dissatisfactions.

Response: We agree with the referee that this is an important issue, but perhaps beyond the scope of a replication study as this is a long debated topic. In spite of this, our response to comment 2 would make some inroads in addressing this issue in the literature.