

RESPONSES TO REVIEWER 1'S COMMENTS

We thank the reviewer for their additional comments. They have spurred us to do a better job of explaining our methods. Below are point-by-point responses to the reviewer's comments.

1. **Comment:** *“The paper is well-written and the various steps in the replication analysis are clearly explained. It represents a good example of a replication analysis that engages with and achieves cooperation from the authors of the original study. It also makes a compelling case that there is no incentive to ‘cherry-pick’ results that confirm or refute those in the original study. This is a useful exercise and there are no obvious concerns about the implementation of the analysis.”*

Response: We thank the reviewer for this positive feedback.

2. **Main points 1:** *“The replication analysis follows the original AKW study in examining three measures of ‘productivity’ – sales, employment and sales per worker – in the models for direct and indirect export shares in Table 1. The original results are successfully reproduced. For direct export shares, a statistically significant (at the 5% level) positive linear effect is obtained only for the sales per worker measure. For indirect export shares, the quadratic term is not significant (at the 5% level) for the sales per worker measure, but is significant for sales and employment. Sales per worker is a standard measure of (labour) productivity, but sales and employment are really measures of ‘size’, not productivity. The differences in the results for the different measures are therefore not entirely surprising. Prediction #1 is explicitly concerned with the effects of changes in productivity not size. In the absence of evidence that sales and employment are reasonably well correlated with productivity, it makes sense to put more weight on the results for labour productivity in the original results and in the extensions testing Prediction #1 in Table 4. The latter results suggest either that effects are very imprecisely estimated (for direct export share) or contradict the theoretically expected signs (for indirect export share). On this interpretation, the results in Table 4 are damaging to AKW’s Prediction #1.”*

Response: We thank the reviewer for highlighting our results on this point regarding Prediction #1. We agree with the reviewer that the results of Table 4, which extend the results of Table 1, do not provide robust evidence of AKW’s Prediction #1. The variable of primary interest - labour productivity – is insignificant for direct export share (panel A) or have wrong theoretically predicted signs for indirect export share (Panel B). The text now emphasizes the results from the labour productivity estimates (see second full paragraph on page 12 of the revised manuscript).

3. **Main points 2:** *“Considering regional sub-samples, as in Table 5, is a reasonable exercise because the model in Table 2 (columns (3.a) and (3.b)) implicitly assumes constant parameters for the relevant population of all intermediary firms. Given that the results for the East region, as with the full-sample results, are consistent with Prediction #2 but those for the Central and West regions are not, the main issue is how to interpret these differences. Economic activity is more concentrated*

in the coastal East region, which, consequently, is much more highly represented in the overall sample. Therefore, it is not surprising that the East sample results are similar to those in Table 2. Other factors are likely to be relevant for the interior regions (beyond product, ownership and destination country fixed effects), but the theoretical model is not intended to consider sub-national geographical factors and the minimal empirical results reported provide no guidance as to whether the model specification is appropriate for these regions. It is hard to judge whether the results for Central and West are to be considered as evidence against the mechanisms in the AKW model or are the product of other types of model misspecification. The latter may also be a possible explanation of the volatility in the estimates for the different years. In addition, it is recognized that the 2005 data are preferable because of the absence of license restrictions. Overall, the implications for Prediction #2 of the differences in the regional and year-by-year estimates are harder to interpret, although the instability in the estimates does cast doubt on any results from the empirical model.”

Response: We take on board the reviewer’s comment that economic activity in China is more concentrated in the eastern region. Hence, it is not surprising that the results of the eastern region dominate the sample and are in-line with the overall sample. To address this point, we provide further analysis in TABLE 5/Panel B by breaking up the eastern region into 3 sub-regions - East1, East2 and East3. These correspond to the Bohai Bay Economic Rim, the Yangtze River Delta Economic Zone, and the Pearl River Delta Economic Zone, respectively. Only the results from East3 provide support for Prediction #2. The other regions produce estimates that are wrong-signed and statistically significant.

4. **Main points 3(a):** *“For the robustness checks for Prediction #3:*

(a) The discussion of regional results in 2 is again relevant. In particular, the theoretical model predicts that higher variable trade costs and higher fixed costs of exporting should affect intermediaries’ share of trade. These costs are measured by home- or destination country proxies at the national level; any additional regional cost factors are therefore not considered. These could be related to additional transportation costs given the vast internal distances involved, distance to the coast, etc. It is not obvious what the effect of omission of such variables would be but it could contribute to variation in regional estimates.

Response: We agree that additional regional factors not accounted for in the model could contribute to the variation in regional estimates between the eastern, central and western provinces of China. To address this concern, we undertake a similar exercise as in TABLE 5 by splitting up the eastern region sample into three smaller sub-samples (East1, East2, East3). While it is likely that the regional differences between the East and the rest of China (Central and West) could lead to these variations, it is less likely that the differences will be large between the three sub-regions of the East.

TABLE 6A provides the results of Prediction #3 for the year 2005. They show that the contrary evidence is not limited to the Central and West regions.

5. **Main points 3(b):** *“For the robustness checks for Prediction #3:*

(b) Some of the proxies are likely to be more relevant than others. Indeed, as noted, the results for distance and GDP are consistent across all model variants, in line with AKW’s Prediction #3. In addition, the online measure of the number of import

documents is held to be the preferred measure and this provides results more supportive of the AKW prediction than the other variants of this variable. Overall, taking into account the various difficulties of measurement and interpretation, it seems to me the results are not as unresponsive of Prediction #3 as an equally weighted ‘success rate’ of statistically significant estimates with expected signs might suggest.”

Response: We agree with the reviewer that some of the results should be weighted more than others. For example, in summarizing the results for Prediction #3, we state: “AKW identify Prediction #3 as their “central prediction”. Here the results are mixed. Overall, 107 out of 175 predictions support the theory. However, these mask an important difference within the five country characteristics examined by AKW. While the test results for distance and GDP strongly support AKW’s theory, those for Chinese population, number of import procedures, and tariffs, do not. Even when we restrict the results to tests where the underlying data are judged to be more reliable (2005), and we use the preferred measure for the number of import procedures variable (Online), the number of successful predictions are never more than half” (page 21).

This is represented in TABLE 8, which is copied below. Under “Re-Analysis and Extension”, we have three columns for Prediction #3. The first column reports all the results (“TABLES 6A + 6B (All)”). The second column only reports estimates using 2005 data (“TABLE 6A”). And the third column only reports estimates using 2005 data where the Online measure of the number of import procedures is used (“TABLE 6A (Online)”).

TABLE 8
Summary of Results from Tests of AKW’s Model

PREDICTION	REPRODUCTION	RE-ANALYSIS AND EXTENSION			
Prediction #1	TABLE 1: 6/9	TABLE 4: 4/9			
Prediction #2	TABLE 2: 3/3	TABLE 5 (All except Panel FE): 3/8		TABLE 5 (Panel FE): 0/1	
Prediction #3 – Total	TABLE 3: 13/14	TABLES 6A+6B (All): 107/175	TABLE 6A: 48/85	TABLE 6A (Online): 4/5	
Prediction #3 – Distance	TABLE 3: 4/4	TABLES 6A+6B (All): 32/35	TABLE 6A: 14/17	TABLE 6A (Online): 5/6	
Prediction #3 – GDP	TABLE 3: 4/4	TABLES 6A+6B (All): 35/35	TABLE 6A: 17/17	TABLE 6A (Online): 6/6	
Prediction #3 – Chinese	TABLE 3: 3/3	TABLES 6A+6B (All): 16/35	TABLE 6A: 5/17	TABLE 6A (Online): 2/6	
Prediction #3 – # Procedures	TABLE 3: 2/2	TABLES 6A+6B (All): 18/35	TABLE 6A: 6/17	TABLE 6A (Online): 3/6	
Prediction #3 - Tariff	TABLE 3: 0/1	TABLES 6A+6B (All): 6/35	TABLE 6A: 6/17	TABLE 6A (Online): 2/6	

Even with the most reliable data (the second and third columns), the estimates for Chinese, # Procedures, and Tariff are not consistent with Prediction #3.

It may be the case that some of the proxies are better than others for testing Prediction #3. However, as AKW did not state that in their article, we don’t feel it is appropriate for us to make that judgment. However, we draw attention to the fact that the theory predicts very well for Distance and GDP. For example, we state in the Conclusion: “*the results reported here should not be interpreted as solely negative.*”

They point to possible avenues for future research. A robust result in our analysis is that the share of exports through intermediaries is positively correlated with distance to trading partner, and negatively correlated with the size of the trading partner's market. This highlights the need for a theoretical explanation for why these two country characteristics in particular should be associated with intermediated trade. We hope that this research stimulates efforts in this direction" (page 22).

6. Minor points

p.1, bottom line: Add 'of' after 'in favor'.

This has been corrected.

p.3, line 5: Tang and Zhang (2012) is not in the reference list.

This reference has been added in the reference list.

p.3, last para, line 1: Insert grave accent over the 'a' in 'Antras'.

This has been corrected.

p.7, para 2: There appears to be a contradiction, or at least some ambiguity, between stating that "while responses are collected at the firm-level, they are only available to researchers at the country-level" and saying that the survey provides proportions of individual firms' sales exported directly and indirectly, as well as various productivity measures. Clearly, the latter is correct.

We eliminated the confusing sentence.

p.9, para 3, line 5: 'large' rather than 'great'? Related to this, given that the sample sizes are often very large, why persist with using conventional 10%, 5% and 1% significance levels to indicate statistical significance, rather than adjusting the significance level to take into account the sample size?

"Great" has been replaced with "large". With respect to statistical significance, we agree that the large sample sizes employed by AKW and us renders the standard significance levels less meaningful. However, as our aim is to stay within the research confines established by AKW, we don't feel comfortable employing a different measure of statistical significance levels.

p.9 and elsewhere in the main text: US spelling is used for 'gray', but UK spelling, 'grey', is used in the table notes.

This has been corrected.

p.12: In the robustness check for prediction #2 that splits the data by geographical region, only the full model's results (corresponding to column 3a in Table 2) are reported. A desire to avoid proliferation of results is appreciated, but do the results for the other (simpler) models yield similar results to those in Table 5? (As it stands, there is no obvious way of selecting a 'best' model from Table 2, as only a goodness

of fit measure is reported, and this is bound to favour model 3 because of the additional regressors.)

The full model is the “best” model as determined by the Akaike Information Criteria. See below for the output:

TABLE 2, Regression (1.b)

Model	Obs	ll (null)	ll (model)	df	AIC	BIC
.	5,193,328	-6817469	-6815130	1	1.36e+07	1.36e+07

TABLE 2, Regression (2.b)

Model	Obs	ll (null)	ll (model)	df	AIC	BIC
.	5,193,328	-6817469	-6802647	5	1.36e+07	1.36e+07

TABLE 2, Regression (3.b)

Model	Obs	ll (null)	ll (model)	df	AIC	BIC
.	5,193,328	-5935972	-5925335	5	1.19e+07	1.19e+07

p.14, fn. 4: Add full stop and closing quote marks.

This has been corrected.

p.20, Table 5: The 2000 result is shaded red, but the estimated coefficient, though having the wrong sign, is not statistically significant at the 5% level. For consistency, this cell should be rose-tinted. Similarly, the statement in the text that “of the remaining three, one is wrongsigned and insignificant (2004), and two are wrong-signed and significant at the 1 percent level (2000 and 2003)” treats the 2000 result as significant at the 5% level. Is the standard error for this result correctly reported?

These errors have been corrected.

p.20, last para: It is claimed that this paper applies Mayo’s ‘severe testing’ approach. But misspecification testing plays a big part in her approach, which is not considered either in the original study or the replication exercises. Therefore, although several useful dimensions of the modelling are probed, it is unlikely that Mayo would consider these sufficient to be considered a severe test.

We agree that Mayo places great weight on specification testing. However, her approach is more than specification testing, as evidenced by her archetypal example of Eddington's 1919 expedition that confirmed Einstein's prediction of the deflection of light by the sun during a total solar eclipse. Given the poor performance of AKW’s model assuming correct specification, we did not see any value in testing further.

p.21, para 3: “... between productivity and labor productivity” – this should refer to productivity and indirect exports.

This has been corrected.

p.21: Finishing on a positive note for future research is good!

Thank you.

p.36, Figure 1: is there a need to define the parameters in the model that appear in the slope term expressions? Either the reader could be referred to AKW's paper or the slope term expressions could be removed.

We eliminated the slope term expressions and now refer the reader to AKW's paper for more detail about the theory.