Online Appendix to:
Does Machine Translation Affect International Trade?
Evidence from a Large Digital Platform

January 4, 2019

Contents:

1 A Simple Theoretical Framework

We provide a simple Nash bargaining framework to illustrate the role of eMT. For simplicity, we model the international markets of a product to be independent of the domestic market of the product. On the demand side, there is a unit-measure of buyers whose highest willingness to pay for the product is w, where $w > 0$. Otherwise buyers differ in the translation cost s, which follows a continuous distribution F.

On the supply side, there is a unit-measure of foreign sellers with zero marginal cost of production, i.e., $c = 0$. Let buyers' Nash bargaining power be α, where $\alpha \in (0, 1)$. Let the market competition be such that buyers capture α of their highest willingness to pay and the price sellers charge, p, is $(1 - \alpha)w$. Then the number of foreign transactions equals the share of buyers who choose to purchase from foreign sellers, or $Pr(w - p - s > 0)$. Given that $w = \frac{p}{1-\alpha}$, the number of foreign transactions can be written as $F(\frac{\alpha p}{1-\alpha})$.

We assume eMT reduces translation cost from s to 0, for simplicity. In this case, all buyers buy from foreign sellers because $w > p > 0$. Then the increase in number of international transactions is $1-F(\frac{\alpha p}{1-\alpha})$. Given this, categories with higher translation costs should expect a larger export response when translation costs are removed. Conversely, categories with low translation costs should not expect as much of an effect of eMT, because international trade in these categories was high even before the introduction of eMT.[1] Proposition 1 formalizes these arguments.

Proposition 1 *Suppose there are two categories with translation cost distributions F and G. The increase in exports due to eMT is higher for the category with translation cost distribution F if F first-order stochastically dominates G.*

Proof. If F first-order stochastically dominates G, then $F(x) \leq G(x)$ for all x. This implies that $1 - F(\frac{\alpha p}{1-\alpha}) \geq 1 - G(\frac{\alpha p}{1-\alpha})$. ■

Next, based on the Nash bargaining framework, the consumer surplus is given by $w-p-s = \frac{\alpha}{1-\alpha}p - s$. If we fix the distribution of translation cost F, consumer surplus of the cheaper product is lower, and therefore buyers were less likely to purchase cheaper products from

[1] Note that these categories could represent different products, number of words in listing titles, and buyers.

foreign sellers before the introduction of eMT. Therefore, the effect of removing s on sales should be greater for the cheaper product than for the more expensive product. The next proposition illustrates how eMT's effect differs between cheap and expensive products:

Proposition 2 *Suppose there are two products with price p_1 and p_2, where $p_1 < p_2$. The increase in exports due to eMT is higher for the product category with p_1.*

Proof. Given $p_1 < p_2$, it follows that $1 - F(\frac{\alpha p_1}{1-\alpha}) \geq 1 - F(\frac{\alpha p_2}{1-\alpha})$. ■

The above theoretical framework leads to a set of testable implications:

- The export increase due to eMT is larger for categories with higher translation costs. In particular,
  - The export increase is larger for differentiated products, because translating the specifics of these products requires higher language skills and hence higher translation costs.
  - The export increase is larger for products with more words in the listing title, because translation costs should increase with the number of words.
  - The export increase is larger for inexperienced buyers on eBay, because these buyers spend less time on eBay and have higher translation costs.
- The export increase due to eMT is larger for cheaper products.

2 Other eMT Rollouts

In this section, we test the robustness of our results using another two rollouts of eMT on eBay. In July 2014, eBay rolled out eMT of item titles and search queries in the European Union to promote intra-EU trade (i.e., items listed on the U.S. sites are not translated). In particular, item titles in English were translated into French, Italian, and Spanish, to facilitate British and Irish exports to France, Italy, and Spain (FRITES). The BLEU score of eMT for English-French translation increased to 48.15, from the BLEU score of Bing at 45.03, and the human acceptance rate (HAR) for French increased from 84.00% to 91.40%. The

BLEU score of English–Italian translation increased from 42.10 to 43.90, and HAR increased from 89.33% to 91.07%. The BLEU score for English–Spanish translation increased from 41.01 to 45.24, and HAR increased from 82.4% to 90.2%.

Additionally, in January 2014, eBay had a rollout of eMT that translated English into Russian to facilitate exports from English-speaking countries to Russia. The BLEU score of eMT for English–Russian translation increased to 23.57, compared to the previous Bing's BLEU score of 14.07. HARs were measured differently for the rollout in Russia back then. In particular, instead of voting yes or no on translation quality, three linguistic experts were asked to rate the quality on a 5-point scale, and averages were taken across the three scores. The resulting average score for eMT is 4.64, compared to 4.00 from the previous translation technology.

We should note that the effects of the Russia rollout of eMT could be contaminated by Russia's annexation of Crimea, which prompted international sanctions. Nevertheless, we think that our continuous DiD approach could partially mitigate the confounding effects from contemporaneous events, assuming that political factors affected exports in a similar way across listings with different title lengths.

We first replicate our continuous DiD estimation on the FRITES and Russia sample. In the FRITES sample, we have three countries (France, Italy, and Spain), 16 word groups, and 18 periods consisting of 9 weeks before and after the policy change. In the Russia sample, we have 1 country (Russia), 16 word groups, and 18 periods consisting of 9 weeks before and after the policy change.[2] As shown in Panel A of Table A1, we estimate a positive per-word policy effect of 0.8% for FRITES and 0.5% for Russia. Ideally, we would like to link differential effects of eMT across languages to differential quality improvement across language pairs. However, this exercise proves to be very difficult in the machine translation community, because BLEU scores cannot be compared across language pairs. That being said, the average improvement in HAR for the three FRITES countries is smaller than that for Latin American countries, consistent with the magnitudes of the estimated changes in exports.

[2]Note that we used only 12 periods (6 weeks before and after eMT's introduction) for analyzing the Latin America rollout, because there were many more observations (18 countries) in that sample.

In Panel B of Table A1, we explore the heterogeneous per-word policy effects of eMT between differentiated products and homogeneous products. For FRITES, we estimate a statistically significant per-word policy effect of 1.2% for differentiated products and 0.3% for homogeneous products. These differential effects remain robust to the inclusion of product and market characteristics. The estimation yields similar results for the Russia sample, and the estimates gain statistical significance with inclusion of additional controls.

In Panel C, we find a larger per-word policy effect for cheaper items than for more expensive items for both FRITES and Russia, with or without additional controls. In Panel D, we estimate a larger per-word policy effect for inexperienced buyers than for experienced buyers for both FRITES and Russia, and this heterogeneity across buyer experience is robust to adding more controls. These effects across product values and across buyer experience are consistent with our main findings in the paper for Spanish-speaking Latin American countries.

3 Export Revenue as Dependent Variable

In the main analyses for Spanish-speaking Latin American countries, we use the logarithm of export quantity as the dependent variable. In this set of robustness checks, we replicate our results using the logarithm of export revenue in dollars as the outcome variable and obtain consistent results.

Specifically, in Table A2 Panel A column (1), we estimate a similar per-word policy effect of eMT on export revenue using our full sample. In Panel B, C, and D, we show that similar heterogeneous policy effects between homogeneous and differentiated products, across product values, and across buyer experience are found in the estimation using export revenue as the dependent variable. Moreover, we find consistent results for all of these analyses as we include additional product and market characteristics in the regressions, as shown in column (2). Lastly, we focus on a shorter time window of six weeks before and after the policy change (i.e., listing title translation), and set up the sample at the weekly level. Columns (3) and (4) present consistent policy effects on export revenue across all specifications.

4 Cross-Country Diff-in-Diff Estimations

In this section, we test the robustness of our results by adopting a standard DiD approach, where we essentially compare the post-policy change in U.S. exports to the treated counties with those to the non-treated countries. We perform the cross-country DiD analysis separately for the three major eMT rollouts: Spanish-speaking Latin America, FRITES, and Russia. It is important to note that the treatment here is the initial eMT event, i.e., the search query translation.

To study eMT's rollout in Spanish-speaking Latin American countries, we take all non-Spanish-speaking Latin American countries that eBay U.S. sellers export to as the control group. As we show in Figure 2b in the main text, the pre-trend assumption holds in the six months before the policy change. In Table A3 column (1), we quantify an overall policy effect of eMT on exports at 17.5% for Spanish-speaking Latin American countries. Note that in this specification, "Post" equals to one after the introduction of eMT for query translation in May 2014; therefore, this is the combined policy effect of introducing eMT for both query and item title translation. In Panel B, we find a larger policy effect for differentiated products than for homogeneous products. In Panel C, we estimate a monotonically-decreasing policy effect in product values for the most part, suggesting a larger reduction in translation cost and hence a greater export increase for cheaper products than for more expensive products. In Panel D, we find that eMT's effect is more pronounced for buyers with less experience. Therefore, the overall policy effect and heterogeneous policy effects are all consistent with what we find in the main cross-title length specifications in the paper.

In the case of FRITES, we use U.K. and Irish exports to non-FRITES countries as the control group for their exports to FRITES countries on eBay. In the case of Russia, we use exports to Russia from non-English-speaking countries as the first control group for exports to Russia from English-speaking countries. Note that there could be spillover effects in cross-country analyses. For example, non-FRITES countries in the control group include Spanish-speaking Latin American countries, which were treated with eMT for U.S. exports. Therefore, we view the results of these cross-country analyses as complementary to our main continuous DiD analyses, rather than as standalone results.

In Table A3 Panel A, we estimate that the introduction of eMT increased U.K. exports to FRITES on eBay by 13.9% in the six months after the policy change. In the case of Russia, we see a 17.1% increase in exports to Russia on eBay in the six months after the policy change. In Panel B, we see significant negative coefficient estimates of the triple interaction terms "Treatment*Post*Homogeneous", indicating that the export increase is larger for differentiated products.[3] In Panel C, for both FRITES and Russia, we find that the export increase is positive for items that are sold for less than $10, and that the increase is smaller or non-existent as the value of a product increases. Lastly, we show that the estimated export increase is smaller for experienced buyers, for both FRITES and Russia. All of these heterogeneous policy effects are consistent with lower translation costs due to eMT, consistent with the results of our main analyses for Latin America using the continuous DiD specification that exploits different title lengths.

[3]In the case of Russia, a negative change in exports for homogeneous products could be a result of political factors, but the sign of this triple interaction term is the same as in the case of Latin American and FRITES countries.

Table A1: Different Countries

	(1)	(2)	(3)	(4)
	France, Italy, Spain		Russia	
	Main Spec	Add'l Controls	Main Spec	Add'l Controls
Panel A. Overall Effect				
No. Words*Post	0.0080	0.0065	0.0054	0.0089
	(0.0009)	(0.0006)	(0.0014)	(0.0015)
Obs	864	864	288	288
Panel B. By Homogeneity of Products				
No. Words*Post	0.0121	0.0094	0.007	0.0103
	(0.0012)	(0.001)	(0.0039)	(0.0034)
No. Words*Post*Homogeous	-0.009	-0.0075	-0.0039	-0.0042
	(0.0036)	(0.0029)	(0.0023)	(0.0024)
Obs	1728	1728	576	576
Panel C. By Product Value				
No. Words*Post	0.011	0.0087	0.0185	0.0177
	(0.0012)	(0.001)	(0.0034)	(0.0021)
No. Words*Post*Value $\in$ [10,50)	0.0007	0.0008	-0.0088	-0.0093
	(0.0016)	(0.0014)	(0.0054)	(0.0039)
No. Words*Post*Value $\in$ [50,200)	-0.003	-0.002	-0.0111	-0.0121
	(0.0016)	(0.0014)	(0.0034)	(0.0029)
No. Words*Post*Value $\geq$ 200	-0.0041	-0.0031	-0.015	-0.0145
	(0.0016)	(0.0013)	(0.0024)	(0.0029)
Obs	3456	3456	1152	1152
Panel D. By Buyer Experience				
No. Words*Post	0.0092	0.007	0.007	0.0103
	(0.0043)	(0.0038)	(0.0032)	(0.0013)
No. Words*Post*Experienced	-0.0055	-0.0062	-0.002	-0.0041
	(0.0035)	(0.0028)	(0.0023)	(0.0017)
Obs	1728	1728	576	576

Notes: We control for variables according to equation (1). In Panel B, we additionally control for the dummy for homogeneous products, its interaction with "No. Words", and its interaction with "Post". In Panel C, we additionally control for the dummies for the four value ranges, their interaction with "No. Words", and their interaction with "Post". In Panel D, we additionally control for the standalone dummy variable "Experienced", its interaction with "No. Words", and its interaction with "Post". Standard errors clustered at the country level.

Table A2: Export Revenue as Dependent Variable

	(1)	(2)	(3)	(4)
	All Data		+/- 6 Weeks	
	Main Spec	Add'l Controls	Main Spec	Add'l Controls
Panel A. Overall Effect				
No. Words*Post	0.0089	0.0129	0.0054	0.0112
	(0.0022)	(0.004)	(0.0014)	(0.0043)
Obs	3024	3024	2592	2592
Panel B. By Homogeneity of Products				
No. Words*Post	0.0141	0.0158	0.0117	0.0134
	(0.0042)	(0.0049)	(0.0048)	(0.0046)
No. Words*Post*Homogeneous	-0.012	-0.0113	-0.102	-0.0071
	(0.0049)	(0.0058)	(0.0052)	(0.0055)
No. Obs	6048	6048	5184	5184
Panel C. By Product Value				
No. Words*Post	0.0121	0.0139	0.0082	0.0117
	(0.0024)	(0.0019)	(0.0011)	(0.0015)
No. Words*Post*Value $\in$ [10,50)	-0.0015	-0.0089	-0.001	-0.0025
	(0.0033)	(0.0027)	(0.0018)	(0.0028)
No. Words*Post*Value $\in$ [50,200)	-0.0023	-0.0028	-0.0013	-0.0033
	(0.0033)	(0.0026)	(0.002)	(0.0025)
No. Words*Post*Value $\geq$ 200	-0.0065	-0.0069	-0.0055	-0.0059
	(0.0032)	(0.0026)	(0.0018)	(0.0027)
No. Obs	12096	12096	10368	10368
Panel D. By Buyer Experience				
No. Words*Post	0.0081	0.0122	0.0081	0.0103
	(0.0022)	(0.0029)	(0.0031)	(0.0035)
No. Words*Post*Homogeneous	-0.0056	-0.0015	-0.0041	-0.0032
	(0.0011)	(0.0014)	(0.0013)	(0.0019)
Obs	6048	6048	5184	5184

Notes: We control for variables according to equation (1). In Panel B, we additionally control for the dummy for homogeneous products, its interaction with "No. Words", and its interaction with "Post". In Panel C, we additionally control for the dummies for the four value ranges, their interaction with "No. Words", and their interaction with "Post". In Panel D, we additionally control for the standalone dummy variable "Experienced", its interaction with "No. Words", and its interaction with "Post". Standard errors clustered at the country level.

Table A3: Cross-Country DiD Estimation

	(1) Latin America	(2) France, Italy, Spain	(3) Russia
Panel A. Overall Effect			
Treatment*Post	0.175	0.139	0.171
	(0.007)	(0.048)	(0.027)
Panel B. By Homogeneity of Products			
Treatment*Post	0.187	0.167	0.164
	(0.016)	(0.018)	(0.035)
Treatment*Post*Homogeneous	-0.062	-0.062	-0.343
	(0.022)	(0.022)	(0.050)
Panel C. By Product Value			
Treatment*Post	0.193	0.209	0.333
	(0.015)	(0.051)	(0.025)
Treatment*Post*Value $\in$ [10,50)	-0.019	-0.031	-0.348
	(0.022)	(0.072)	(0.036)
Treatment*Post*Value $\in$ [50,200)	-0.056	-0.013	-0.377
	(0.022)	(0.072)	(0.036)
Treatment*Post*Value $\geq$ 200	-0.059	-0.120	-0.391
	(0.022)	(0.072)	(0.036)
Panel D. By Buyer Experience			
Treatment*Post	0.176	0.187	0.201
	(0.014)	(0.045)	(0.048)
Treatment*Post*Experienced	-0.061	-0.087	-0.171
	(0.019)	(0.049)	(0.080)

Notes: We control for variables according to equation (1). In Panel B, we additionally control for the dummy for homogeneous products. In Panel C, we additionally control for the dummies for the four value ranges. In Panel D, we additionally control for the standalone dummy variable "Experienced". Standard errors clustered at the country level.