**Online Appendix: Response to Previous Referee Reports**

Note: Our Responses to previous referees are in Dark Blue. We thank all of our previous referees for taking the time to provide feedback, and for improving our paper.

Reviewer #3: I am keeping this review short so that I can get it back to the editor and author quickly.

**Response**: Thank you! Indeed, you were very fast, which we appreciate.

I have a number of concerns about this paper as a possible publication at the EER.

1. This area is over-researched. There are already 2 meta-analyses (Rose and Havranek) and 2 lengthy reviews (Baldwin and Santos Silva and Tenreyro’s Currency unions in prospect and retrospect. 2010 Annual Review in Economics which is not cited\*).

**Response**: We actually agree that this area is over-researched, particularly since Campbell (2013) showed that the original Currency Union Effect was driven by endogeneity, omitted variables, and CU switches with missing data. And yet, very good journals, such as the JIE and EER, continue to publish papers which find large CU effects, albeit without acknowledging the critique in Campbell (2013). More than 40 papers a year are written on this topic. The point of our paper is that this literature should end, or rather, that it never have started to begin with. By not publishing our paper, you are leaving the door open to additional wasted resources being devoted to this topic.

Both of the meta-analyses are quite dated. While I believe meta-analyses are interesting, and can tell us something, not all studies included in the meta-analyses are of equal value. Many of the studies in both the Rose and Havranek meta analysis have clear flaws, such as including CU switches coterminous with warfare and/or communist takeovers, or not including a UK\*“Ever UK colony” time trend to control for the gradual decay of UK colonial trade. Second, Campbell (2013) was not a part of either meta-analysis. Third, these meta-analyses are quite dated at this point. Fourth, we come to drastically different conclusions than either of the meta-analyses.

We agree that Baldwin had a very nice overview of the literature up to 2005. His review left many reasons to doubt the very large estimates of Rose, but in the end he left the puzzle open and concluded, based on a gut feeling rather than regression results, that a CU effect of 5-15% is plausible. The paper does not articulate a smoking gun (or several smoking guns) about what is driving the result like Campbell (2013) or this paper do for the extended data set. Also, we, like Campbell, cast doubt on an effect as large as 5-15%, which in our subjective view is still a large trade impact, on par with estimates of the Smoot-Hawley tariff, for example.

We apologize for not citing Santos Silva and Tenreyro (2010). Probably, it deserves a citation, although the focus of that paper is not just on the CU effect, and they run a total of three regressions on a short panel (1993-2007) for the Euro only that omit country-pair fixed effects, and do not plot pre-treatment trends or control for trends. However, there are easily more than 200 papers in this literature, most of which contribute something. We do cite 44 of them. Thus, we’d like to take this moment to apologize to all future referees on this topic – it’s quite likely that, even if you’ve written on this topic, we might not have room to cite you. Another one of the comments that we have gotten is that our paper is already too long.

1. There are already several prominent papers challenging Rose so I do not think scholars in the field are unaware of the concerns over Rose’s claims that CUs have very large effects.

**Response**: To clarify, we actually wrote this paper in part because we were tired of seeing papers presented on this topic at conferences, and published in journals which we would consider ourselves very lucky to publish in (the JIE and EER), which were in fact **not at all aware** of the key problems in the original CU effect literature, such as the need for trends, dynamic controls, or the need to plot pre-treatment trends. (Just have a quick look at the last 3-4 years of European Trade Study Group conference papers.) Very good journals continue to publish clearly flawed papers finding large CU effects. This suggests that the specific problems with the CU effect literature are, in fact, not very widely known. The most prominent critique of Rose was Baldwin’s excellent overview. However, he still argued for a fairly large effect and effectively left the puzzle open.

Also, I think you are confused with what our argument is. We find there is no support for claims of significant trade effects, not that just the very large estimates are over-blown. We believe that it is not well known that the evidence for *any positive* effect is actually quite weak, and sensitive to specification. As evidence for this claim, below in your report, ***you yourself argue for a positive CU effect***. We would like to know why? Which part of the evidence we presented, or presented in Campbell (2013), did you not find convincing? Inquiring minds would like to know.

1. The endogeneity issue is probably the key issue on which new ideas are needed but I don’t see a major new insight in this paper.

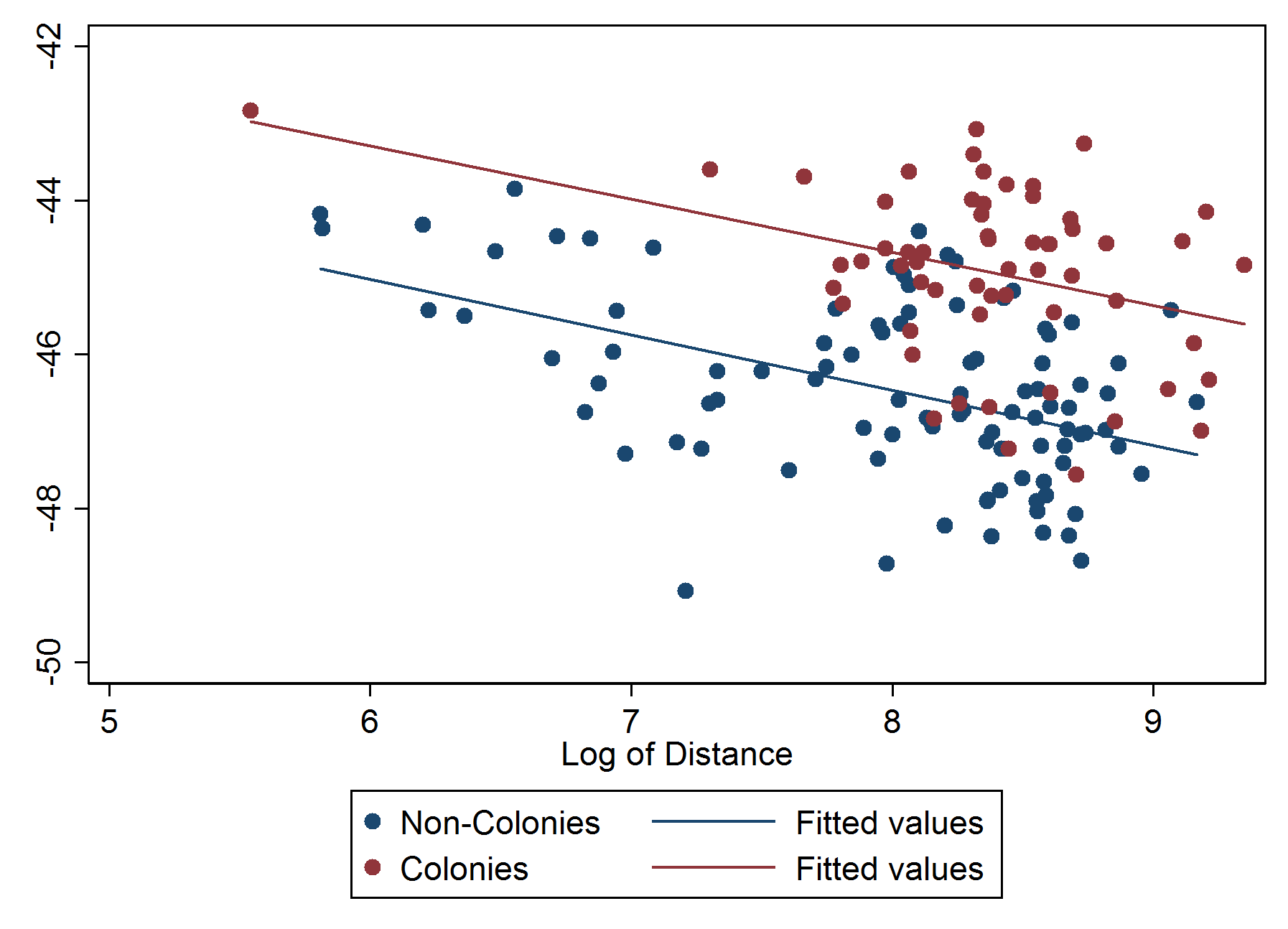
**Response**: The insight is that, even with this much larger dataset, the results in GR (2016) (and also in papers such as Saia, 2017) are not robust. We cannot reliably say that there is a measurable CU effect greater than zero. We think this is valuable since it shuts down a large and still rapidly-growing literature, a literature that you seem to agree is already too large. As mentioned previously, the problems highlighted in Campbell (2013) are also not well known in the first place.

1. Specification also matters and it appears this paper exlusively uses linear in logs. There was almost no mention of this issue and yet SST and de Sousa both obtain quite different results when using PPML.

**Response**: This is a fair point. In the current draft, we also add the PPML in the appendix. We decided not to use the PPML previously primarily for two reasons: (1) as we mention in our paper, in the first draft of the GR (2016) paper, they also showed that they get vastly different results using a PPML. They just concluded that there was something strange about the PPML estimator. Thus, it was already known that using the PPML generates different results, which, in our view, is another reason to discount the CU effect. However, the point we want to make is that the answer to the CU effect does not lie in whether you prefer the linear in logs or the PPML estimator. Even using linear in logs, there’s not a strong case for a positive CU Effect. (2) At first, we could not solve the computational problem of including all the necessary FEs. Now we can, thanks to the algorithm of Larch, Wanner, Yotov, and Zylkin. But, since these guys also do a PPML, and show wildly different results, we decided to put the PPML in the appendix. The evidence using the PPML mostly confirms the basic result that the positive impacts of CUs on trade are not robust.

There are several other reasons to discount the PPML method. First, the central argument for it, according to Santos Silva and Tenreyro (2006), is the existence of a large number of true zeros in the trade database. However, in fact, we believe that there are no true zeros, or almost no true zeros, in an aggregate trade database. This is because of travel and entrepot trade. Benin and Guatemala may not trade with each other directly in the trade statistics, until a Guatemalan in New York City takes a cab from a Benin uber driver, a trade transaction that won’t show up in the official statistics. Similarly, Slovenia and Ghana may not trade directly, but they both trade a bunch with Belgium, a trade entrepot. And many zeros in our trade data are sandwiched in between much larger values, suggesting they are simply missing observations. A PPML estimator using zeros puts much more weight on missing/erroneous data.

Secondly, Santos Silva and Tenreyro (2006) show that gravity results are different with a PPML, but not necessarily better. They find, for example, that colonies do not increase trade. They argue that this is the true value, and that OLS is simply flawed. However, simple scatter plots of data (below) suggest otherwise. The graph is simply a plot of log trade/GDP vs. distance for the UK, with regression lines run separately for colonies vs. non-colonies. It suggests that UK colonies trade much more than non-colonies. Only 3 red dots (colonies) even lie below the blue regression line, while only two blue dots lie above the red regression line, which is perhaps surprising given that the UK has free trade with all other EU countries. One wonders if the reason is that former colonies were more likely to have missing data/(erroneous) zeros.



Finally, as we discuss in the paper and as Campbell (2013) notes, there happens to be a lot of missing data directly before/after CU switches. Campbell (2013) found that dropping the CU switch pairs coterminous with missing data happened to weaken the CU effect greatly, although he could kill the result even without resorting to this. We believe this data is better dropped, although note that the results do not end up hinging on this.

Finally, let us just note that our goal here is not to wade into the PPML vs. Log-linear debate, nor do we necessarily believe the log-linear version is superior. The more compelling reason we do log-linear is because this is what GR (2016) use, and so we can show that even using their preferred methodology that the results are not robust, much less with PPML.

1. The main finding here is that adding the geopolitical controls drives the CU coefficient down to about 0.1. It seems possible that CU effects are identified by a small number of cases and that if we controlled for events surrounding those cases the effect would lose significance.

**Response**: To clarify, we do not find significant CU effects.

But I’m concerned that one could get rid of a \*true\* effect by judiciously adding controls that have the right timing. So the question is what arguments justify the presence of this particular set of non-standard controls?

**Response**: Sure, it is possible that we have cherry-picked which observations to throw out, and that we have done some reverse p-hacking. This is probably a more salient critique of Campbell (2013), as he throws out nearly one-fifth of the CU switches in GR (2002). However, if you look at the list of 26 switches coterminous with warfare or other major geopolitical events – which include ethnic cleansing episodes, communist takeovers, and coup d’etats -- I think you’ll agree that these were all major events likely to overshadow changes in CU status. (Any disagreements?) We only drop a very small fraction of the data points for the reason of warfare – 25/423 switches with time series variation – why doesn’t the CU effect show up on the other 398 switches? To be fair, there are a lot more switches with missing data, but excluding these is not necessary to kill the CU effect.

Our approach is to evaluate each CU relative to an intuitive control group, and thus our controls merely reflect this. Thus, we compare Western European countries who joined the Euro to those that did not. We do the same for the EU countries in WE. We find that using all of Western Europe vs. the WE EU countries doesn’t matter – one still won’t arrive at a positive and significant Euro effect. For Eastern European Euro members, we simply include as a control a bunch of other nearby countries that did not join the Euro. Which control groups/controls, specifically, did you find objectionable?

Also, we’d like to stress that when we plot the pre- and post-treatment effects, even without controls, the evidence suggests that the CU switch was not the cause of the CU effect. Thus, to some extent, the controls are not even necessary. What we show is that when we include the controls, the trends go away. But so do the effects. Controlling for country-pair time trends also kills the results, absent controls. This is a “death by a thousand blows” paper. It doesn’t hinge on one control.

1. T-stats are not the key issue. The true average treatment effect of currency unions should be be positive. At least I can’t think of a mechanism by which using the same currency would lower trade.

**Response**: Perhaps it “should be positive”, but, intuitively, I also thing the magnitude should be much smaller than the standard errors we are measuring (typically around 8%). This makes it likely that any measured positive impact will simply be noise. I am pessimistic in part because the key feature of a CU here is a 1:1 par. Why should a 1:1 par foster trade but not a 1.2:1 par? It’s not intuitive why we should assume a 1.2:1 par does not foster trade, but that a 1:1 par does.

In the Euro case, it isn’t too hard to imagine how the Euro might have lowered trade, albeit indirectly. Some economists believe the Euro has been terrible for the economies of southern Europe, and thus, since GDP has collapsed in countries like Greece, so too has all of its international trade, including with other Euro countries. Trade is pro-cyclical, and varies more than GDP, so even trade/GDP should fall.

The debate is about magnitudes, not significance. Yet this paper does not seem to reach a bottom line on the preferred estimate.

**Response**: I agree that t-scores aren’t the key issue, but they are important. If I’m a policy-maker in Athens, Rome, Lisbon, or Madrid, managing economies in deep depressions looking for a way out, then I would care about t-scores. If I’m told joining the Euro for a certain increased trade of 7%, and with 99.99% confidence that it increased trade significantly, then that is a tangible reason to stay in. If, on the other hand, I’m told that in fact there is a 40% chance there was in fact no measurable trade effect, I’m going to discount this evidence. The t-score here is policy-relevant.

The bottom line of this paper is that the significant positive estimates of the CU effect are not robust, and thus that policy makers in places like Athens should discount a large CU effect in their calculus of whether to stay in the Euro.

1. The before-after plots in Figure 12 and 13 are the way I would think about this. The issue of pre-trends deserve more attention in text and Figures.

**Response**: Good point. This issue rightly deserves more emphasis. We have tried to emphasize this more in the text.

1. A short comment on the fragility of Glick and Rose’s most recent results might be worth looking at for EER. This 31 page paper before appendices is overkill.

**Response**: We have tried to tighten things up. Now we finish in less than 28 pages before References. Note that we have 8 Tables and 10 Figures, which take up over half of the space. The actual writing is almost certainly less than 15 pages.

1. The expositional style will rub some readers the wrong way. Examples “These insights paved the way for Campbell (2013)” sounds pretentious.

**Response**: Perhaps you are correct. But the intention was to credit the earlier work, and to discount Campbell (2013). Perhaps it came across as pretentious rather than humble, as it was intended. We’ve changed the language to “Campbell (2013) drew on these insights…” The intention here is to give credit to this earlier literature, not to Campbell.

“However, Rose would typically respond with larger data sets,” seems to belittle or besmirch Rose’s efforts.

**Response**: None of the critiques in Campbell (2013), which themselves were not necessarily new, were addressed by Glick and Rose’s papers in (2015), or (2016a) or (2016b). This despite the fact that the authors thanked Campbell for sharing his (2013) paper to them (although declined to cite). The new aspect of these papers was that they used even larger datasets. Of course, *any criticism* of other research will fundamentally belittle or besmirch it. I disagree with your fundamental point that all criticism – even when correct and justified -- should be eliminated from academic papers.

Comments on “Breaking Badly: The Currency Union Effect on Trade”

Summary

This paper assesses the robustness of the Glick and Rose (EER 2016) finding of large currency union (CU) effects. When the authors change the sample, employ multiway clustering, and add controls, the CU effect largely disappears.

General

The topic is important and, as the authors recount, there is a longstanding controversy on

the impact of CUs on trade. I am not convinced the results disappear until I understand the controls used in specifications with exporter-year, importer-year, and bilateral fixed effects. These controls are important in knowcking down the CU effect (see point 2 below).

The paper should be much shorter and to the point.

Specific comments

1. Specification: I would only report specifications with bilateral, exporter-year, and importer year fixed effects. This is clearly preferred to a specification without country-year FEs since unobserved country-specific influences are captured. That will save a lot of space. There is no reason to discuss results of foawed specifications that do not employ the full set of fixed effects.

**Response:** First of all, thank you very much for taking time to referee our paper.OK, we took out some of those, but we do think it is instructive to offer an apples-to-apples comparison withthe prior literature, and the only real way to do that is with the CPFE model. Showing that the results also don’t hold in this model shows that the impact of CUs on trade is not driven by using a different regression model/the FEs. Lastly, it isn’t that it is necessarily the “wrong” model, it is essentially just a model with fewer controls. It is a way of showing robustness of results as you add in more FEs. That the results tend to jump around when more FEs are added is a way we know that they are not particularly robust. It is standard in empirical economics to display regression tables in which subsequently more controls are added. This is similar.

2. I do not understand what controls are added to specifications with bilateral, exporter-

year, and importer-year fixed effects.

* Table 3: “Columns (2) and (5) add EU\*year and EE-Euro Area\*Year interactive

FEs."

* Table 7: “In column (4), we add in a number of intuitive controls meant to capture the various factors discussed in the previous section. These include (a) UK Colony\*year interactive fixed effects, (b) Ever UK colony\*year FEs, (c) year\*Ever EU effects.

How can you add region-year FEs when there are country-year FEs? They will be collinear. Are they bilateral-year? If so, are they absorbing important CU effects? (For example, a bilateral EU-year FEs for 1998 and 2004 absorb most of the CU effect.) The addition of these controls dramatically reduce the estimated CU effect and it is unclear to me what they are.

**Response:** Region-year FEs are not collinear with country-year FEs, since the data is bilateral.Consider the case of Portugal. The country-year FEs in the regression will be a Portugal\*Any country\*Year FE for each year and each observation with Portugal as an importer (and another set for Portugal as an exporter). Then we add in another set of dummies for each year in which the importer and exporter are both EU countries. The set of Portugal-EU country pairings, which include Portugal-France, Portugal-Spain, Portugal-Great Britain, etc., will only be a subset of these FEs, and only a subset of the Portugal\*Any Country\*year FE pairings, since the set of EU countries do not comprise every country in our dataset.

Also note that the EU is not the same as the Euro. Sweden, Denmark, and Great Britain are in the EU, but not in the Euro. In addition, Norway, Switzerland and Iceland are in Western Europe but not the EU. So, there are over 100 observations per year of countries not in the Euro which are effectively in the control group.

In Eastern Europe, most countries are actually not in the Euro, so there the Eastern Europe Euro\*Euro Area\*year sample will actually be smaller than the sample of EE Euro countries.

Intuitively, what these controls do is tell us how much the Euro increased trade relative to trade between EU countries, or trade between all countries in Western Europe. *I.e.*, it is about selecting the proper control group, and a dynamic one at that.

Admittedly, this was probably tough to understand, and so in our new draft, we hopefully make this easier to understand.

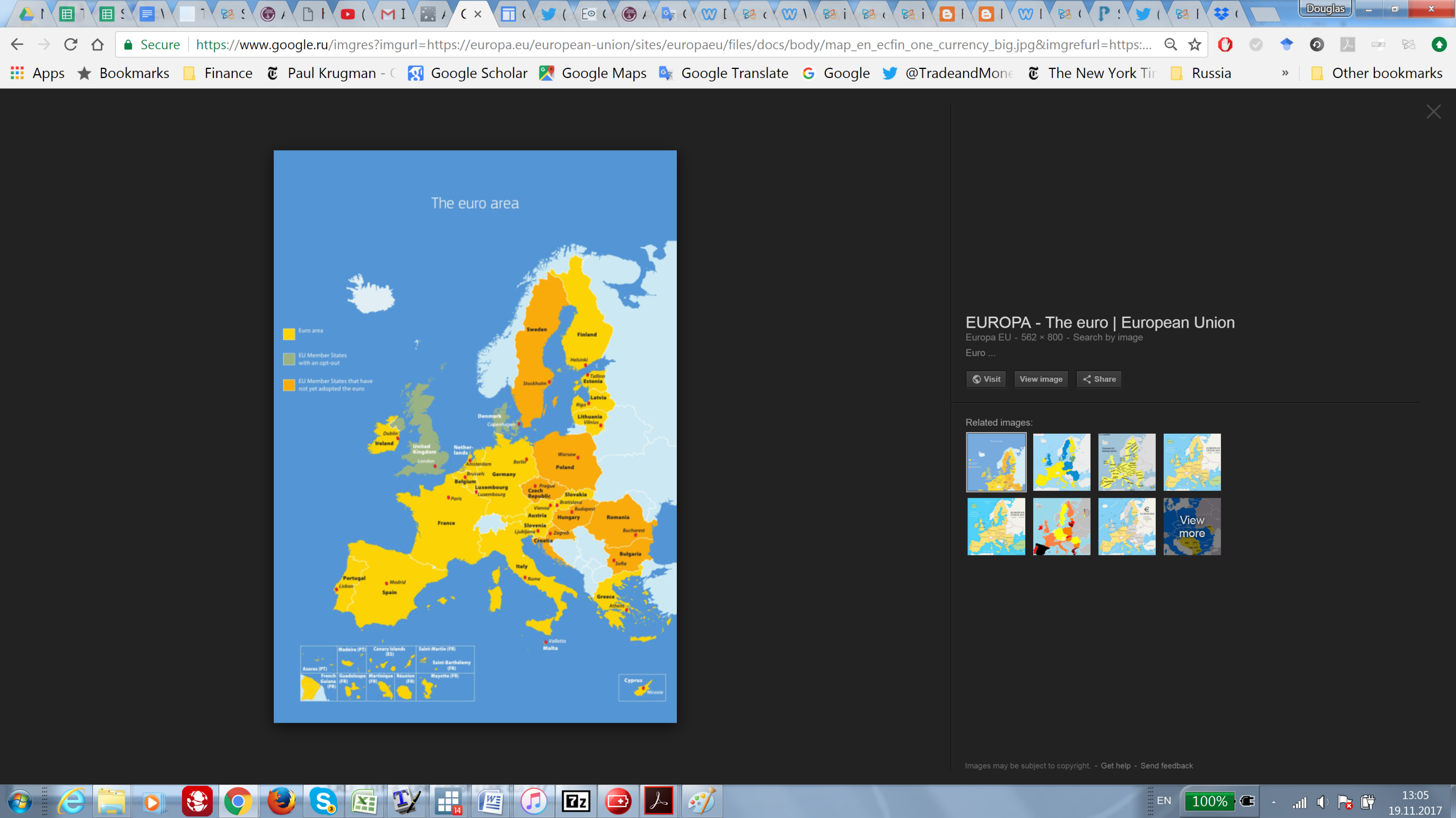
3. Control countries: Much of the paper examines specific CUs to “help us find appropriate control groups, similar to the “matching" approach which has been tried by others." Given a specification with country-year and bilateral fixed effects, it is not clear how varying the set of countries included in the regression affects the CU estimates and what is the preferred sample. There must be a much better discussion of how the sample influences results in a structural gravity specification.

**Response:** One small note here. I think the “gravity” element of this is less relevant than it appears, given that we have country-pair FEs which will wash out the distance element. The research design is to compare the evolution of trade between Euro Area countries, and other similar countries which are not on the Euro (but in the EU or Western Europe), before and after the formation of the EU.

* When the sample is reduced, there may be a small number of control countries, thereby reducing the power of the test. What countries in the EU that are not in Western Europe?

**Response:** Surprising as it may seem, much of Eastern Europe is now in the EU. See graph below. The Baltics, Poland, the Czech Republic, and Romania are all examples of Eastern European/former communist countries not in the Euro.

There are also three countries in Western Europe which are not in the EU, and still other countries in Western Europe in the EU but not the Euro zone. There are six countries in Western Europe that do not use the Euro – this is plenty large enough to form a control group. In Eastern Europe, there are at least a dozen countries that do not use the Euro.



I also feel too much time is devoted to the large number of “case studies”.

**Response:** Perhaps. We have now shortened this section slightly. However, we feel that, in this literature, too much time and stress has been placed on the large panel data regressions in which the authors have no idea what is going on with each individual CU. When you plot the CU effects over time for each individual CU, you see that in ***no cases*** does the time series evidence support a large treatment effect. We believe this is, in fact, powerful evidence against a large effect. From these, we can see that the large effects are driven by pre-treatment trends, events 25 years after a CU begins, or disappear with intuitive changes in the control groups.

Even Reuven Glick, after being told of this paper, offered the defense that they just had so much data, that they tried to put in the standard controls, and just very hard to do more than that when you are sitting on one million observations. In such as case, it’s hard to keep track of specific events, such as communist revolutions or the war between India and Pakistan, which happened to coincide with CU dissolutions.

Also, elsewhere in your review you complain that you can’t understand the controls, and don’t understand the motivation. But, in part, the point of this section is to motivate and explain the controls used in the panel regression portion of the paper.

4. I like Figure 12 (Impact of CU Exits and Entrances, GR Specification) as it seems like pre-trends explain the results. However, footnote 12 states, “[GR 2016] should further be commended for plotting pre- and post-treatment trends, and for adopting a new specification with one-directional exports as the dependent variable while controlling for importer\*year and exporter\*year interactive fixed effects." Don't they observe the same trends shown in Figure 12?

**Response:** Well, while we can’t read the minds of Glick and Rose, we can observe that their version of Figure 12 does have problematic pre-treatment trends. We interpret this as key evidence that their treatments are not random. At some point in their paper, they argue that the Euro likely had anticipation effects. However, these anticipation effects start in the early 1950s, long before the Euro was ever conceived.

5. I'd like to see Table 7 appear very early.

**Response**: We can understand this. However, we do worry that we need time and space first to motivate our controls. Indeed, you also asked for a better motivation and explanation of the controls in the panel regression. This we include in the section immediately preceeding the panel regression, with the discussion of individual CUs.

6. What about zeros? Other estimators (PPMLE)?

**Response**: We now show (in the appendix) that one gets wildly different results using the PPML. There are some reasons not to use the PPML (see above), which include that we’d rather not include zeros, which we don’t believe are true zeros, but more likely missing data.

7. I am not keen of the “Breaking Badly" part of the title.

**Response**: Well, there is a long history in this literature of playfull titles, *e.g*., “Honey I Shrunk the Currency Union Effect on Trade”. In this case, we’d like to point out that it is a triple entendre, as many currency union dissolutions

**Comments from the Editor:**

I have heard from two reviewers and find there is insufficient support to ask for a  
revision. This is not because the paper is not interesting or wrong, but because it is not novel enough and leaves too many questions unanswered. Since we can accept only about 8% of submissions, we hand out invitations to revise only when we have solid evidence that the paper has a clear chance to make a mark in terms of novelty to meet the bar of the the referees and the EER. Both referees are leaders in the field and both have questions regarding the novelty. Both referees also have questions regarding the methodology (e.g, see the issue of zeros and fixed effects). Having worked in the area myself, I do appreciate the anecdotes but they themselves do not make for an entirely novel insight since the problems with the GR finding are well known - indeed they have been the origin of a thriving literature. For your guidance, the reviewers' comments are included below.

**Response**: Thank you very much for taking time to review my paper. It may be that the problems with the Glick and Rose (2002, 2016) are well-known in some circles, but I can’t help but noting that just last year, the European Economic Review published Rose’s (2016) paper, which repeated the problems pointed out in Campbell (2013).

As we wrote above “To clarify, we actually wrote this paper in part because we were tired of seeing papers presented on this topic at conferences, and published in journals which we would consider ourselves very lucky to publish in (the JIE and EER), which were in fact **not at all aware** of the key problems in the original CU effect literature. (Just have a quick look at the last 3-4 years of European Trade Study Group conference papers.) Very good journals continue to publish clearly flawed papers finding large CU effects. This suggests that the specific problems with the CU effect literature are, in fact, not very widely known. The most prominent critique of Rose was Baldwin’s excellent overview. However, he still argued for a fairly large effect and effectively left the puzzle open.”