Many unions, one estimate? Disaggregating the currency union effect on trade

Joseph Kopecky[a]*

[a] *Department of Economics, Trinity College Dublin, Dublin, Ireland*

* jkopecky@tcd.ie

Many unions, one estimate? Disaggregating the currency union effect on trade

A large literature purports to estimate the impact of currency unions on trade. Often ignored in these estimates are the dramatic differences in the characteristics of countries adopting common currencies, with most estimating either an aggregate currency union effect, or grouping all non-Euro arrangements together and treating the Eurozone separately. Further, when estimating standard gravity equations of trade, there is little work considering issues of selection into currency unions, ignoring the extreme differences in observable characteristics that not only exist between different types of unions, but also between currency union and non-union country-pairs. In this paper, I show that estimating the effects of individual currency unions on trade implies a wide range of coefficients. Using inverse propensity score weighted methods, I show that adjusting these disaggregated gravity equation estimates to account for the substantial differences between currency union pairs and their non-union counterparts also meaningfully impacts the estimated policy effects.

Keywords: currency unions; trade; propensity scores; gravity

Subject classification codes:

1: Introduction

"Happy families are all alike; every unhappy family is unhappy in its own way."

- Leo Tolstoy, Anna Karenina

The currency union effect on trade is a contentious topic, with revised attention in recent years now that data on the European Monetary Union (EMU) has a long enough time series to provide estimates of its effect. In this paper, I explore two often ignored aspects of this literature. The first is that estimates of currency union trade impact generally fail to account for the imbalance that exists between treatment and control samples, with the non-union members having substantial observable differences from countries in a currency union, making them likely a poor comparison group. My evidence suggests that this may account for the variability in currency union estimates, which have been shown in recent work on the EMU by Rose (2017) to depend predominantly on the choice of sample. Second, looking at individual currency unions reveals substantial heterogeneity, not only in the estimates of currency union effects across disaggregated unions, but also in the nature of this imbalance between members of these groups and non-currency union observations. Most currency unions studied in large trade datasets are developing countries using either a colonizer's currency, that of another large, developed nation, or who are in multilateral arrangements with other developing economies with their common currency pegged to a large country. I find evidence that modelling the choice of entry into a disaggregated currency union, and that using propensity score methods to create a closer comparison group to members within that union, substantially changes the estimates of the trade impacts of these policies. My findings suggest that one-size-fits-all measures of these policies are likely quite poor and provide misleading evidence of their impact on bilateral trade. While trade improvement is not the only factor in determining whether a currency union is an optimal[1] policy choice, it reflects one of the large potential benefits. Improving and better contextualizing these estimates will better inform these monumental policy decisions.

Since the seminal work of Rose (2000), a great deal of effort has gone into the estimation of the currency union effect on trade. Work such as Nitsch (2002) purported to have shrunk the eye-popping estimates from the original work which were by the

[1] See Alesina, Barro, and Tenreyro (2002), Bayoumi and Eichengreen (1993), and more recently Aizenman (2016) and Chari, Dovis, and Kehoe (2020) for a discussion on the many costs and benefits of entering into a monetary union.

author's own admission "*embarrassingly and implausibly large*," (Rose, 2002). These estimates were smaller, but still substantial when utilizing the time-series variation among currency unions in Glick and Rose (2002). With a growing time series in one of the largest currency union policy changes in history, a large body of research suggests that the effect for the EMU looks smaller. Polák (2019) conducts a meta-analysis of this literature, studying 3323 estimates of the euro's trade effect with careful controls for methodology and design of the study. These suggest a range somewhere between 2% and 6%. Glick and Rose (2016) estimate the EMU effect as larger, arguing for a roughly 50% increase in trade among members. Rose (2017) suggests that the reason behind the diminished EMU effect in other research comes from limiting the sample to include only large, rich, economies. My analysis will confirm this, while showing that properly weighting and truncating the sample to account for comparability between currency union and non-union members does change estimates, likely improving them.

Less well established are the effects of non-EMU currency unions. It is somewhat understandable that the creation of such a large multilateral currency union arrangement would create such interest, but much can be learned from the other currency unions in the Glick and Rose (2016) sample, which for the most part are either small developing countries using a larger trading partner's currency, or a group of relatively small countries adopting a common currency. In their paper studying the EMU, Glick and Rose (2016) provide estimates for disaggregated unions, showing the effect of their gravity equation estimates for a number of other currency unions. Their estimates suggest dramatic differences among these groups, though this is not their focus. Saiki (2005) shows that for developing countries who use a larger country currency that the currency union effect is asymmetric. Their work finds that exports from the USA and France to dollarized and CFA Franc countries increase in response to the policy, but that the effects are neutral on imports into those larger countries from their developing counterparts. Campbell and Chentsov (2017) consider the disaggregated currency unions used in my sample (and others), showing that the historical context matters for these policies. In similar work, Campbell (2013) showed that much of the aggregate currency union effect on trade is driven by the major geopolitical events causing currency union dis-allusion, largely decolonization and war. There is evidence that such bias is likely present as Head, Mayer, and Ries (2010) show that after decolonization trade falls substantially between a country and its former colonizer, while Glick and Taylor (2010) show that wars have large trade impacts. Time series estimates that do not account for some dynamic factors will show this to be a positive effect of the currency union, which picks up the colonizer relationship despite the fact that the correlation is likely spurious. Campbell (2013) shows that estimates of the aggregate currency union relationship like those estimated in Glick and Rose (2002) do not hold up to accounting for these events with dynamic controls.

Adopting the language of the causal inference literature, consider currency unions as the *treated* group. Non-currency union observations are thus a *control*. If country pairs were randomly assigned into treated and control groups, then one could simply take a weighted average of the estimates to arrive at an average treatment effect. Of course, this is not how these relationships are formed. Controlling for various observable characteristics and fixed effects can help improve estimates given the *selection on observable* assumption, but still falls far short of causal inference in cases where selection into the policy treatment is endogenously determined. I will use two different estimators which weight the estimates from gravity equations with probabilities of treatment derived from estimations of selection into currency unions. The inverse probability-weighted regression adjustment (IPWRA) estimator, described in detail in Imbens (2004), Lunceford and Davidian (2004), and Wooldridge (2007), simply incorporates these weights into a standard gravity equation. The augmented inverse probability weights AIPW, described in detail in Glynn and Quinn (2010), augments an inverse probability weighted (IPW) estimator of group means with regression adjustments. These estimators seek to resolve to problems with previous estimates of the currency union effect on trade. First, they will work to reduce problems of poorly matched samples within and outside of currency union pairs. Second, they will show that each disaggregated currency union has dramatically different models of selection, suggesting that aggregate estimates of currency unions likely provide flawed understanding of policy treatment effects.

One of the many critiques levelled against existing aggregate gravity equation estimates[2] of the currency union effect on trade, is that they fail to account for the endogeneity of choice of currency unions. Countries enter into these arrangements as a result of factors that will bias estimates of their trade relationship. Some existing work has attempted to address these selection problems. Persson (2001) does this on the original aggregate currency union estimate from Rose (2000), finding smaller implied estimates. Chintrakarn (2008) carries out a similar exercise on the EMU, using matching estimators to show a reduced trade effect. In both cases using estimates of probability of entering the union to create matched treatment and control groups weakens the currency union estimate. Rebuttals point out that these estimators throw out the majority of data, often choosing one or a few observations to use as comparisons for each observation of a country-pair in a currency union. Matching estimators such as

---

[2] For a full discussion of this debate in the context of the Eurozone estimates see Frankel (2010).

those of Persson (2001) and Chintrakarn (2008) use similar logic to my approach. However, the IPWRA and AIPW estimators I use are better able to balance the potential usefulness of the well-studied gravity equation approach with these probability weighted estimators. In one of the few studies to look at individual currency unions in detail, Campbell and Chentsov (2017), find that currency union estimates are highly sensitive to controls and that most existing estimates appear to be driven by spurious relationships that become apparent when considering the historical context, or in some cases simply plotting the trade relationships. While their work does not specifically use matching[3], their findings strongly motivate the need to better contextualize the situations of developing country currency unions. My estimates will be a simple data-driven version of such an approach.

My estimation method also shares something in common with estimates from Barro and Tenreyro (2007), who use an IV estimation motivated by the model of Alesina and Barro (2002), whereby probit estimations are used to estimate the probability that a country adopts the currency of one of six potential anchor countries. Their approach is particularly concerned with exactly the selection issue that mine seeks to address. They calculate the probability that two countries adopt a common anchor as an instrument in a bilateral trade regression, finding quite large estimates in line with Rose (2000). My estimation strategy is a more reduced form approach that makes fewer assumptions about the reasoning behind currency union membership. Also unlike theirs, I will focus on the disaggregated currency union effect, while they estimate one effect across many currency unions. It is hard to directly compare to their results, as they do

3 In the case of countries using the Australian Dollar they do something of an ad hoc version of this approach with New Zealand trade relationships as a matched control.

not incorporate the time-varying country fixed effects suggested by Head and Mayer (2014), nor the dyadic pair effects which substantially decrease the trade effect in Glick and Rose (2002) and are shown by Baldwin and Taglioni (2007) to be important empirically in estimation of currency union effects. However, my estimates suggest that an update of their work, which focuses specifically on the motivations of unions who adopt common currencies to peg against a larger country,[4] would be useful on a disaggregated basis.

In section 2, I describe the data and methods used. Here I provide further motivation by showing that union and non-union pairs have large differences in observable characteristics and define the estimators I will use to resolve these differences. I present estimates for standard unweighted gravity equations, the IPWRA on a full and truncated sample, and the AIPW estimators in section 3. section 4 contextualizes these findings, provides some discussion of paths for future work, and concludes.

2. Data and Methodology

My estimations use the International Monetary Fund's *Direction of Trade* data. For ease of comparison, I will use exactly the same coverage as Glick and Rose (2016), who's study contains bilateral trade data from 200 countries from the time period of 1948 to 2013. Before discussing estimation methods, I establish some motivating evidence that the observable characteristics of countries in currency unions are quite different from non-currency unions. Moreover, these differences change depending on the currency union relationship at hand, suggesting that not only is it important to carry out some rebalancing adjustment, but to also do so separately for individual disaggregated

---

[4] The CFA Franc and East Caribbean currency union are both of this type.

currency unions.

I consider the largest seven currency unions in this sample, as measured by the number of country-pair-year observations. I limit to these as restricting my outcome model of trade to theory consistent methods includes the use of a large number of fixed effects, which at times makes estimation of unions with few observations infeasible. These are: the European Monetary Union (EMU), the CFA Franc, the East Caribbean currency union (ECCU), the Indian Rupee zone, and countries using Australian dollars, US dollars, and British pounds. The purpose of this study is to explore the extent to which directly modelling the individual currency relationship affects estimates of trade. Restricting my estimates to the Glick and Rose (2016) data improves comparisons to well-known estimates, but it is likely that this limits the ability of my propensity score model to fully contextualize the many reasons why currency unions form and break. However, it will become clear that even a fairly stylized model of selection is able to substantially improve the comparability of currency union and non-union members, with large implications for their estimated effects on trade.

Table 1 shows the mean and standard deviation for four variables under policy treatment (currency union) and control (non-currency union) groups. The first two of these are standard gravity equation variables, the log product of the bilateral pair GDP, and GDP per capita. The second two reflect the absolute value of the difference between these two variables in exporter and importer countries. These latter values are of particular interest as they distinguish between trade partners of relatively similar size and income (such as in developing-to-developing pairs), or those with large imbalances in these characteristics. I do not report means for the full sample as they would be in general indistinguishable from the control sets, which can be seen to vary only trivially across these different currency union groups. This is because even the aggregate currency union only consists about 4.2% of the overall sample, so the non-treated observations make up the vast majority of total observations. The bottom of the table reports the difference in treated and control sub-populations. Large positive or negative deviations here mean that the currency union in question has large observable differences in characteristics from the country-pairs that are being used as a comparison in standard estimates.

Table 1: Summary Statistics: Disaggregated Currency Unions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All CUs	EMU	CFA	ECCU	Aussie Dollar	British Pound	Rupee	US Dollar
Control (CU=0)								
$log(GDP_1 GDP_2)$	49.49	49.44	49.49	49.47	49.45	49.45	49.45	49.45
	(2.641)	(2.681)	(2.667)	(2.667)	(2.686)	(2.687)	(2.687)	(2.686)
$log(GDPpc_1 GDPpc_2)$	17.47	17.44	17.48	17.45	17.45	17.45	17.45	17.45
	(1.780)	(1.799)	(1.788)	(1.807)	(1.806)	(1.806)	(1.805)	(1.806)
$log(\|GDP_1 - GDP_2\|)$	25.71	25.68	25.71	25.69	25.68	25.68	25.68	25.68
	(1.903)	(1.941)	(1.918)	(1.928)	(1.941)	(1.941)	(1.940)	(1.938)
$log(\|GDPpc_1 - GDPpc_2\|)$	8.920	8.898	8.918	8.900	8.899	8.899	8.900	8.899
	(1.422)	(1.441)	(1.423)	(1.440)	(1.440)	(1.441)	(1.439)	(1.441)
Treated (CU=1)								
$log(GDP_1 GDP_2)$	47.15	52.79	45.15	40.76	45.84	49.68	49.30	52.26
	(4.149)	(2.213)	(1.236)	(1.418)	(1.018)	(1.415)	(1.242)	(1.652)
$log(GDPpc_1 GDPpc_2)$	16.56	20.57	14.28	17.79	18.37	17.30	14.11	18.89
	(2.875)	(0.706)	(1.105)	(0.878)	(0.480)	(1.140)	(1.210)	(1.452)
$log(\|GDP_1 - GDP_2\|)$	23.80	26.84	22.10	19.82	26.88	27.09	27.06	29.46
	(3.029)	(1.431)	(1.328)	(1.386)	(0.431)	(0.165)	(0.992)	(0.542)
$log(\|GDPpc_1 - GDPpc_2\|)$	7.575	9.107	6.487	7.970	10.07	8.823	5.271	9.425
	(1.903)	(1.318)	(1.531)	(1.094)	(0.350)	(0.522)	(1.779)	(1.073)
$Tr - Con$: lGDP	-2.34	3.34	-4.33	-8.70	-3.62	0.23	-0.15	2.81
$Tr - Con$: lGDPpc	-0.91	3.13	-3.19	0.34	0.92	-0.15	-3.34	1.44
$Tr - Con$: lDiff. GDP	-1.91	1.16	-3.61	-5.87	1.20	1.41	1.38	3.79
$Tr - Con$: lDiff. GDPpc	-1.34	0.21	-2.43	-0.93	1.17	-0.08	-3.63	0.53
N	738876	738876	738876	738876	738876	738876	738876	738876

Consider the two multilateral arrangements of developing countries: the CFA franc and East Caribbean Currency Union (ECCU). These are monetary unions of small developing countries who not only share a common currency, but also peg that currency to a large, developed country (the French Franc/Euro and United States dollar). These pairs have substantially smaller output relative to the non-currency union sample, with the ECCU countries having roughly average living standards. Moreover, these groups are unique among the rest of the currency unions in the sample as they have smaller average difference in both output and living standards. This is true for the aggregate currency union, but not for any other disaggregated member (though a few others have smaller differences in per-capita output). The CFA and ECCU unions reflect country-pairs that are relatively homogeneous in terms of size and average income compared to not only the control set, but other currency union pairs.

The Eurozone represents pairs of countries that are much larger and richer than the average in the sample, reflected by the large positive deviations from their product GDP and GDP per-capita. While they are relatively diverse in terms of size, with a larger difference within the pair in terms of GDP than in the control set, they have somewhat close to average differences in GDP per-capita. So, while the EMU sub-sample is quite rich relative to the controls the differences among trading partners in terms of level of richness is somewhat similar. I keep the EMU as one unit here but splitting out original members and recent entrants might exaggerate some of these differences, with the original members being an even richer and more homogeneous group on these macroeconomic measures.

The last four unions in this table are a group that all share the characteristic of being pairs consisting of a very large and very small economy. This is reflected in all four having a positive difference across treated and controls in the absolute value of the difference between trading partner GDP. They have little else in common, with the sheer size of US and UK GDP making their product of GDPs larger than the control groups, despite most partner countries being extremely small, while the Australian dollar reflects much smaller joint size. The difference in living standards in both Australian and US dollar arrangements is substantial, reflecting much larger gaps in living standards between the large central country and those adopting their currency than among the average trading partners in the sample, while the opposite is true for the Indian Rupee zone, where per-capita output in India was quite low relative to its smaller trading partners.

There should be two broad takeaways from Table 1. The first is that currency union countries are very different from the average non-union trading partner in this sample. The second is that every currency union member is different in its own way, with some common themes that can be identified, but no systematic pattern that holds across groups. It is clear that context matters a great deal, and looking at this table one would not expect an aggregate estimate of currency union to do a good job describing the experiences across these groups.

2.1. Baseline Currency Union Estimator: The Gravity Equation of Trade

I specify the theory consistent representation of the gravity equation of trade, as described in Head and Mayer (2014). This involves including a full set of exporter-year and importeryear fixed effects to my estimation. These time varying fixed effects should pick up some of the spurious correlations revealed to bias prior estimates in Campbell (2013) but are unlikely to solve all problems of identification and endogeneity. In addition to this, I will follow Glick and Rose (2016), whose preferred estimates include pair specific fixed effects. I wish to keep my gravity specification close to theirs so that my estimates will be easily compared. I estimate the following:

$$\ln(X_{ijt}) = \gamma CU_{ijt} + \beta Z_{ijt} + \lambda_{it} + \psi_{jt} + \varphi_{ij} + e_{ijt} \tag{1}$$

Where X_{ijt} are exports from country i to country j at time t, CU_{ijt} is a dummy variable representing a currency union relationship between the country-pair in year t, Z_{ijt} is an arbitrary set of time-varying controls, λ_{it} exporter-time fixed effects, ψ_{jt} importer-time fixed effects, and φ_{ij} time-invariant country-pair fixed effects. This specification of the gravity equation is identical to the preferred estimates in the section of Glick and Rose (2016) that uses this "newer" theory consistent export model. I will present baseline estimates that are quite close to theirs. Additionally, I will include their regional trade agreement variable in Z_{ijt}, though will omit other common gravity controls as these are generally captured by the rich set of fixed effects.

One difference between my baseline specification and that of Glick and Rose (2016) is that I use currency unions that are *non-transitive*. This is actually the definition in their earlier work (Glick and Rose, 2002). Transitivity suggests that if countries x and y are in a currency union, while countries x and z are also in a currency union, then countries y and z are also in one. While it is true that this transitive property accurately reflects a shared currency among countries, it has implications for how I will estimate my first stage selection into currency unions. Multilateral currency union arrangements, such as the EMU, will inherently have such transitive properties as all countries in the group join together through multilateral agreement. This transitive property comes into play in countries who unilaterally adopt a particular currency, such as developing countries using dollars, who then find such *sibling* relationships with other countries who make the same choice. Since the choice to join a currency union *together* is not necessarily made in these cases I choose to ignore these observations, as my later propensity score methods will attempt to model that selection directly. They are dropped from analysis entirely so as to not introduce bias through the control set, though there is little impact on my baseline estimates relative to those of Glick and Rose (2016) who includes them and including them in my matching estimates appears to not have substantial impact.

2.2. Propensity Score Weighting: Two Doubly Robust Estimators

While it is possible that the multilateral resistance terms, along with my control for regional trade agreements may properly account for problems of selection in Equation 1, the large gap in Table 1 of currency union and non-union populations warrants some caution. I use inverse propensity score weighting in the estimation of Equation 1 to adjust the regression estimates on the currency union effect of trade in an attempt to properly balance the characteristics of currency union and non-currency union estimates. In particular, I leverage two forms of regression adjustment that have the *doubly robust* property.[5] These estimators' model both the outcome variable of interest, here log exports, as well as the selection into treatment groups (currency union membership). The doubly robust property refers to the fact that these estimators should provide consistent estimates of the average treatment effect if *either* of these two models are correctly specified. For more detailed derivation of these propensity score estimators in a macroeconomic context see Jordà and Taylor (2016).[6]

---

[5] For detailed discussion on these and other propensity score estimators see: Imbens (2004), Lunceford and Davidian (2004), and Wooldridge (2007).

[6] Though their estimators use dynamic local projections framework these can easily be converted to a static context as I do here.

There is a long history of propensity scores in the literature on currency unions and trade, where Persson (2001) used matching methods to determine comparable groups among broad currency union membership to show that under matching techniques the estimates of currency union impacts on trade are much lower. Chintrakarn (2008) uses similar methods on the Euro area, again finding much smaller estimates than those using an unweighted gravity equation. Matching methods inherently throw out a large amount of data by choosing to either use only the best matches, or a small sub-sample of close matches. This is because they are designed to seek identification by attempting to convert observational data to something closer to a randomized control trial, through re-randomization. To my knowledge there is limited work using these regression adjustments with propensity score data in the trade literature. Given that the gravity equation of trade in Equation 1 is a well studied and theoretically grounded object, it seems reasonable to assume that the information from this regression adjustment is useful, even if there are gains to be made through the traditional propensity score adjustments.

To the extent that treatment and control samples are significantly different, it is possible to improve non-weighted estimates of Equation 1 by implementing an inverse propensity score weighing with regression adjustment (IPWRA). I follow similar notation to Jordà and Taylor (2016), who provide a full discussion of how these doubly robust estimators are derived in a potential outcomes framework in a macroeconomic context. The IPWRA estimator is:

$$\widehat{ATE}_{IPWRA} = \frac{1}{n_1^*}\sum\left[\frac{CU_{ijt}\left(m_1(Z_{ijt},\hat{\gamma})\right)}{\hat{p}_{ijt}}\right] - \frac{1}{n_0^*}\sum\left[\frac{(1-CU_{ijt})\left(m_0(Z_{ijt},\hat{\gamma})\right)}{1-\hat{p}_{ijt}}\right] \tag{2}$$

Where CU_{ijt} is a dummy representing whether a country-pair is in a currency union time t, $\hat{p}_{ijt}$ is the predicted probability of treatment, estimated from some first-stage model of likelihood of currency union membership. The term $m_{0/1}(Z_{ijt}, \hat{\gamma})$ is an estimate of the conditional mean control/treated groups given a set of controls,[7] Z_{ijt}. This conditional mean is the estimation coming from Equation 1, with $\hat{\gamma}$ the vector of estimated regression coefficients. In principle it is possible to allow for separate estimates of $\hat{\gamma}$ among treated and control populations, but because the number of disaggregated currency union observations is relatively small this is not possible in practice due to the rich set of multilateral resistance terms and dyadic fixed effects. Allowing different conditional means on treated and control would require estimating a weighted version of Equation 1 on just the treated subpopulation, and doing so as described would generally imply more parameters to estimate than observations. As such, in practice I must assume that the coefficients estimating these conditional means are the same across these groups. Finally I normalize the inverse probability weights with $n_1^* = \sum \frac{CU}{\hat{p}}$ and $n_0^* = \sum \frac{1-CU}{1-\hat{p}}$ as suggested in Hirano and Imbens (2001) and Imbens (2004), which simply requires the weights to sum to one within each group. My second estimator is the AIPW, which is given by:

[7] To condense notation I refer to the control set as simply $Z_{ij,t}$, but this includes the rich set of fixed effects in Equation 1.

$$\widehat{ATE}_{AIPW} = \frac{1}{n}\sum \left\{ \left[\frac{CU_{ijt}\left(m_1(Z_{ijt}, \hat{\gamma})\right)}{\hat{p}_{ijt}} - \frac{(1 - CU_{ijt})\left(m_0(Z_{ijt}, \hat{\gamma})\right)}{1 - \hat{p}_{ijt}} \right] - \frac{CU_{ijt}(\ln{(X_{ijt})}}{\hat{p}_{ijt}(1 - \hat{p}_{ijt})} \left[(1 - \hat{p}_{ijt}) m_1(Z_{ijt}, \hat{\gamma}) + \hat{p}_{ijt} m_0(Z_{ijt}, \hat{\gamma}) \right] \right\} \tag{3}$$

The first term in this expression is a standard inverse probability weights (IPW) estimator, which simply takes a weighted difference of the group means across treated and control populations. The second term is an adjustment which is mean zero in large samples. A particularly useful property of this adjustment term is that it provides stability of the estimator when probability weights approach the zero/one extremes, a problem that has been shown to introduce instability in estimates of the IPWRA estimator. These properties make the AIPW estimator particularly useful in the case of trade data, where sample sizes are quite large and can take advantage of some of these benefits, but large fractions of the data in the control group are assigned weights near zero. However, they also put more weight on the IPW first-stage model, which in this case is less well theoretically justified. Work such as Glynn and Quinn (2010) show using a Monti Carlo design that this estimator, IPW estimators with a regression adjustment term, preforms dramatically better than standard IPW, but does not compare it directly to the properties of the IPWRA. As with the IPWRA I follow the baseline estimates from Jordà and Taylor (2016), who in their primary estimates specify the AIPW with an assumption of equal coefficients for treatment and control populations (ie, identical $\hat{\gamma}$ terms) in estimating the conditional mean. This, as above, is necessary in nearly all disaggregated currency unions due to lack of sufficient observations when estimating Equation 1 on the treated population alone.

2.3. Modeling Selection Into Currency Unions

Both the IPWRA and AIPW estimators require a first-stage estimation of the likelihood of entering into currency unions for each bilateral pair over time. My first stage specification to predict these probabilities of treatment with a logit model that uses some standard gravity estimators (the log product of GDP and GDP per capita). In addition to I include terms that aim to fit differences among trading partners.

$$CU_{ijt} = \theta_0 + \theta_1 ln(Y_{it} \times Y_{jt}) + \theta_2 ln(y_{it} \times y_{jt}) + \theta_3 ln(Dist_{ijt}) + \theta_4 ln\left|Y_{it} - Y_{jt}\right| + \theta_5 ln\left|y_{it} - y_{jt}\right| + \theta_6 \left|g_{it} - g_{jt}\right| \quad (4)$$

In Equation 4, in addition to the first three terms, which are standard gravity equation estimates of size of output ($Y_{i/j,t}$), incomes ($y_{i/j,t}$), and distances ($Dist_{ij,t}$) between the two countries, I also include the differences in output, per-capita GDP, and GDP growth rates ($g_{i/j}$). These are important to my motivation because the standard gravity terms may do a poor job of capturing differences between a multilateral union of rich countries (like the EMU) and matches that take place between wealth countries and their developing trade partners. I also include two terms for the square of the log product of GDP and GDP per-capita. This improves fit of the model, and Millimet and Tchernis (2009) suggest that there are benefits of *over-specifying* the propensity score estimator. Because I wish to keep my results as comparable to the existing literature on currency unions and trade, I work only with variables readily available in standard bilateral trade datasets. Given that work such as Rose (2017) shows that choice of data has large implications on estimates, it seems prudent to work from the same starting point to demonstrate the estimates obtained from the IPWRA and AIPW estimators.

As my results below will suggest, estimations of Equation 4 perform quite differently in terms of capturing this selection process across the disaggregated currency unions I study. Work such as Campbell and Chentsov (2017) describe the historical context behind changes that occur in currency union membership. An ideal study might attempt to incorporate these kinds of historical and geopolitical shifts into this full trade data, including other potential controls such as demographics or institutional/structural shifts. This might mean specifying quite different versions of Equation 4 depending on the context relevant to a particular group. Improving the model in Equation 4 to include various demographic, institutional, and political controls might provide further insight into some of the important differences between these unions and identify a more appropriate control group of trade partners. As such I view my results that follow as a first step to ground these disaggregated effects in the literature, rather than a definitive estimate of any given policy effect.

2.4. Dealing With Limited Overlap

While my first stage estimates from Equation 4 generally provide good fit, with large amounts of overlap between the treated and control sub-populations, many countries-pair observations are assigned with incredibly small probabilities of entering into a currency union. This can has been shown to be a source of bias in propensity score estimates. It is common in the literature on propensity scores to trim the sample, with somewhat arbitrary rule of thumb conventions quite common. Imbens (2004) suggests a number of solutions to lack of proper overlap. The first is to trim the sample to exclude outliers, whose potential influence increases in small samples. Export trade datasets are much larger than many of the observational studies where conventional cut-offs of [0.05, 0.95] or [0.1, 0.9] are common. However the number of extreme outliers in many of my estimations are still large, and potentially a cause for concern. As such, I will present estimates of the IPWRA estimator on both the full sample, as well as those where any estimated probabilities of treatment fall outside of the [0.001, 0.999] range. In most cases no currency union observations are dropped, but the majority of data, which falls below this range is removed. This is because for some disaggregated measures of currency unions, very few trading partners in the control set are close comparisons.

One contribution of this paper is to contribute to the discussion of why currency union estimates vary so greatly. Rose and Stanley (2005), conducts a meta-analysis on the early estimates of currency unions on trade finding, among hundreds of estimates, large positive effects, that imply a doubling of trade. Later estimates that focus on the EMU tend to be much smaller, with effects in the range of a 2%-6% increase in trade. Rose (2017) conducts a meta-analysis, investigating *why* such a large amount of variation in estimates for the EMU exist and concludes that there are systematic differences based on sample choice. His conclusion is that using larger samples tends to increase the estimates of the currency union on trade, while smaller samples shrink them. He summarizes his findings, saying: *"truncating the sample by omitting countries that are small or poor biases downward the estimates of the country-time fixed effects ... this leads to downward bias in the estimated partial effect of EMU on exports"* (Rose, 2017). He refers to dropping countries from the sample as *inappropriate*, without providing clear justification why keeping the data is inherently better. I will show that the non-currency union pairs have substantial differences with those in a currency union, a problem that can become even larger when disaggregating the currency unions. Weighting on propensity scores without changing the estimation sample can slightly improve this metric, but large differences remain. Trimming the sample to remove extreme outliers, while keeping propensity score weights appear do a better job creating an appropriate control set.

The claim that more data is better need not be true for its own sake, particularly when the data is expanded by introducing observations with poor comparison to those who are receiving the policy treatment. Consider the analogy of an observational medical study where the population receiving a drug is systematically more likely to be sick than the general population, as is usually the case. Including large amounts of *untreated* observations can easily make the drug look like it's causing poor health outcomes. An obvious choice might be to construct a control sample of those individuals with similar medical history, pre-existing conditions, and diet. This is precisely what propensity score estimators attempt to do in situations where you cannot pre-randomize and must try to do so ex-post. While these re-randomizations do not improve credibility of estimation results to the extent that a randomized control trial would,[8] they certainly are a step in the right direction from the perspective of rebalancing the sample to avoid these more obviously poor comparison groups.

3. Results

I now present results for my estimation of Equation 1 under various weighting specifications. I begin by simply estimating a standard unweighted gravity equation to set a baseline of my estimates relative to the existing literature. I will then present estimates using propensity score weights using the IPWRA estimator, including those where the full sample is kept and those where the sample is trimmed to remove the

[8] There are still likely unobserved confounders that are not being addressed.

potential bias induced by large outliers. Finally I show estimates of the AIPW, finding that these estimators imply in some cases quite different results.

3.1. Baseline Estimates: Disaggregated Gravity

Here I show estimates that are quite similar to those estimated with the same *theory consistent* method of estimating the gravity relationship via exports as in Glick and Rose (2016). My sample will be slightly smaller than that of Glick and Rose (2016), due to missing data for some of the controls used in the logistic model of currency union selection. While these data can be used in these estimations, which do not include these weights, I instead present estimates on a matched sample with the weighted currency union estimates below.[9] This sample reduction ensures proper comparability between my estimates using propensity score weights with these unweighted currency union effects. The differences in estimates in the first two columns with their equivalents in Glick and Rose (2016) are quite small.

The estimates in Table 2 are not surprising, given that they match the existing Glick and Rose (2016) disaggregated results. They reflect a substantial estimated currency union trade impact, with the average effect of 0.39 reflecting a 48% ($e^{0.39} - 1$) improvement in exports upon entering into a currency union. These of course vary a great deal with estimates slightly larger, but quantitatively similar estimates to this baseline in the EMU, Aussie Dollar, and British pound currency unions, whose estimates reflect a 56%, 52%, and 60% increase respectively. Much larger are the estimates for the CFA Franc and Indian Rupee with implied increases of 112% and 73%. Puzzlingly low, but in line with the results from Glick and Rose (2016) is the

---

[9] I am able to perfectly replicate the Glick and Rose (2016) results when using the full data.

incredibly large negative value for the ECCU, with a point estimate reflecting an 80% reduction in trade.

Table 2: Disaggregated Gravity: Unweighted Baseline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Currency Union	0.39***								
	(0.03)								
EMU		0.47***	0.45***						
		(0.04)	(0.04)						
CFA Franc		0.75***		0.75***					
		(0.08)		(0.08)					
East Caribbean CU		-1.60***			-1.61***				
		(0.19)			(0.19)				
Aussie Dollar		0.41†				0.42			
		(0.26)				(0.26)			
British Pound		0.48***					0.47***		
		(0.08)					(0.08)		
Indian Rupee		0.58***						0.55***	
		(0.17)						(0.17)	
US Dollar		0.12							0.13
		(0.11)							(0.11)
Exporter-Year FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Importer-Year FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Dyadic FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
R2	0.861	0.861	0.862	0.860	0.860	0.860	0.860	0.860	0.860
N	732143	732143	723300	726606	721847	720271	721277	720923	721385

*Standard errors clustered by country-pairs, † $p<0.15$, * $p < 0.10$, &** $p < 0.05$,***_$p < 0.01$.*

3.2. Estimating Selection into Currency Unions

I now turn to my estimates of inverse probability-weights. Before estimating Equation 2 or Equation 3, I must first estimate a first-stage model of the probability that any country-pair observation will be engaged in a currency union. These estimated probabilities from these models will be used in all of the IPWRA and AIPW estimators that follow to create the inverse propensity score weights for each disaggregated currency union. I estimate Equation 4 using a logistic regression.[10] I separately estimate these probabilities using identical models for each of the disaggregated currency unions. Acquiring separate estimates of these currency unions, rather than fitting a single first stage model for aggregate currency union treatment is important to my approach. The particular types of countries in these various currency union pairs look quite different from each other based on their observable characteristics in Table 1. In an attempt to improve on the validity of the treatment and control groups it is important to model these separately. In Table 3, I show the coefficients from this logistic estimation for the full aggregated currency union as well as for each of the seven disaggregated unions I study.

Though the coefficients themselves provide little information on my estimators of interest, there are a few remarks worth making about Table 3. The first is that using a pseudo-R^2 measure of fit there is a wide range of performance in terms of variability in selection into these currency union groups picked up by this estimator. Using the aggregate currency union is quite poor, but so too is the fit for countries using the British pound. This is likely a measure of the variability of countries within these samples with the British pound union, which covers a wide range of diverse countries from the start of the sample until the late 1960s, and now (other than the UK itself) consists of just a few small territories who still use the currency. The East Caribbean currency union on the other hand is a small group of quite similar countries, and the

---

[10] Using a probit model instead has little impact on the results, but appears to have slightly worse fit in most cases as judged by pseudo-R^2.

model does a good job fitting treatment, though as will be apparent when truncating the sample below, finds few non-ECCU trade partners that make suitable comparisons in the data. Another notable feature of Table 3 is that the coefficients differ significantly across sample, further driving home the need to estimate selection separately for these various currency unions.

Table 3: First Stage Logistic Regression

	All CUs	EMU	CFA	ECCU	Aussie Dollar	British Pound	Rupee	US Dollar
Log(|GDP1-GDP2|)	-0.05***	0.14***	-0.40***	-1.10***	1.61***	0.92***	2.26***	2.81***
	(0.01)	(0.02)	(0.01)	(0.07)	(0.08)	(0.04)	(0.13)	(0.07)
Log(|GDPpc1-GDPpc2|)	-0.11***	-0.13***	-0.03*	-0.17*	0.77***	-0.28***	-0.59***	-0.22***
	(0.01)	(0.02)	(0.01)	(0.07)	(0.12)	(0.03)	(0.05)	(0.04)
Log(Dist)	-1.00***	-1.11***	-0.97***	-2.70***	-1.50***	-0.80***	-2.38***	-1.81***
	(0.01)	(0.02)	(0.02)	(0.12)	(0.11)	(0.05)	(0.13)	(0.06)
Log(|GDP1GDP2|)	-3.04***	1.16***	23.29***	-6.16***	42.60***	15.72***	23.61***	8.37***
	(0.06)	(0.31)	(0.62)	(1.19)	(5.02)	(1.58)	(3.92)	(1.21)
Log(|GDPpc1GDPpc2|)	-3.51***	18.44***	9.22***	24.52***	53.93***	8.04***	-2.86**	2.02**
	(0.06)	(1.20)	(0.28)	(2.78)	(7.82)	(0.83)	(1.05)	(0.70)
|growth1-growth2|	-2.14***	-9.37***	-1.28***	-0.54	3.90***	-0.50	-2.99	1.74**
	(0.18)	(0.83)	(0.21)	(1.58)	(0.92)	(0.72)	(1.67)	(0.61)
Log(|GDP1GDP2|)²	0.03***	-0.01***	-0.26***	0.06***	-0.48***	-0.16***	-0.25***	-0.08***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.05)	(0.02)	(0.04)	(0.01)
Log(|GDPpc1GDPpc2|)²	0.10***	-0.42***	-0.32***	-0.65***	-1.45***	-0.23***	0.05	-0.05**
	(0.00)	(0.03)	(0.01)	(0.08)	(0.21)	(0.02)	(0.03)	(0.02)
Observations	733356	724511	727816	723059	722170	722486	722134	722596
Pseudo R-squared	0.294	0.447	0.502	0.920	0.531	0.128	0.555	0.472

*Standard errors clustered by country-pairs, † $p<0.15$, * $p < 0.10$, &** $p < 0.05$, ***_$p < 0.01$.*

3.3. Currency Union and Trade: IPWRA Estimates

With first stage estimates in hand I can now estimate my IPWRA specification. I estimate Equation 2 by using the probability weights from the relevant currency union estimation in Table 3 to weight regression estimations with a specification described in Equation 1. Since only one set of weights can be applied at a time I estimate only the individual estimates, rather than the joint estimation from column two of Table 2.

Because I estimate each disaggregated currency union individually I omit all other currency union pairs from this specification[11] so as to not taint my control set with observations that are in a different currency union relationship. I first look at IPWRA estimates that retain the full sample, corresponding to the same samples as those estimated in the standard gravity estimations above. These estimates are shown in Table 4:

Table 4: Disaggregated Gravity: IPWRA Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Currency Union	0.43***							
	(0.02)							
EMU		0.40***						
		(0.02)						
CFA Franc			0.50***					
			(0.10)					
East Caribbean CU				-1.41***				
				(0.08)				
Aussie Dollar					0.32**			
					(0.13)			
British Pound						0.43***		
						(0.03)		
Indian Rupee							0.51***	
							(0.09)	
US Dollar								0.05†
								(0.03)
Exporter-Year FEs	✓	✓	✓	✓	✓	✓	✓	✓
Importer-Year FEs	✓	✓	✓	✓	✓	✓	✓	✓
Dyadic FEs	✓	✓	✓	✓	✓	✓	✓	✓
R2	0.921	0.924	0.873	0.889	0.887	0.938	0.881	0.934
N	732143	723300	726606	721847	720271	721277	720923	721385

*Standard errors clustered by country-pairs, † p<0.15, * p < 0.10, &** p < 0.05,***_p < 0.01.*

At first glance these look similar to the estimates from the unweighted regression. All estimates of individual currency unions are smaller (in absolute value)

[11] Including, as mentioned above, any currency unions from transitive relationships.

than those in Table 2, with the aggregate currency union estimate slightly larger. With the exception of the two outliers in the ECCU and US dollar estimates are now much more tightly clustered around the average estimate. The comparable estimate from Glick and Rose (2016), which finds a 0.43 coefficient on the EMU trade impact holds up quite well to including propensity score weights in this way. A logical first question to ask is: given this weighting scheme, are treated an control populations better matches for one another? The inverse propensity weights can be applied to the summary statistics presented in Table 1, to see if conditioning on these weights brings populations in balance. These are presented in Table 5.

Table 5: Summary Statistics: Weighted on Full Sample

	All CUs	EMU	CFA	ECCU	Aussie Dollar	British Pound	Rupee	US Dollar
Control (CU=0)								
$log(GDP_1 GDP_2)$	47.67	49.52	49.47	49.50	49.50	49.51	49.51	49.51
	(3.859)	(2.641)	(2.659)	(2.639)	(2.625)	(2.635)	(2.636)	(2.636)
$log(GDPpc_1 GDPpc_2)$	16.61	17.49	17.45	17.47	17.47	17.47	17.47	17.48
	(2.879)	(1.786)	(1.798)	(1.779)	(1.776)	(1.778)	(1.779)	(1.779)
$log(\|GDP_1 - GDP_2\|)$	24.18	25.72	25.69	25.72	25.72	25.72	25.72	25.72
	(2.429)	(1.901)	(1.934)	(1.902)	(1.898)	(1.901)	(1.901)	(1.903)
$log(\|GDPpc_1 - GDPpc_2\|)$	7.672	8.925	8.903	8.924	8.925	8.924	8.924	8.925
	(2.016)	(1.421)	(1.440)	(1.421)	(1.420)	(1.421)	(1.422)	(1.421)
Treated (CU=1)								
$log(GDP_1 GDP_2)$	44.39	49.58	46.05	44.00	46.14	50.43	47.75	52.31
	(4.607)	(2.068)	(1.306)	(0.594)	(0.647)	(1.472)	(0.665)	(2.586)
$log(GDPpc_1 GDPpc_2)$	16.29	19.39	16.31	17.09	18.01	17.48	14.44	17.95
	(2.885)	(0.730)	(1.972)	(0.442)	(0.767)	(1.273)	(0.379)	(1.166)
$log(\|GDP_1 - GDP_2\|)$	21.68	25.44	22.60	22.58	26.59	27.07	26.47	29.19
	(2.900)	(1.811)	(0.843)	(0.496)	(0.369)	(0.149)	(0.314)	(0.538)
$log(\|GDPpc_1 - GDPpc_2\|)$	7.051	9.366	7.757	8.949	9.883	8.736	7.516	9.825
	(1.887)	(0.917)	(1.546)	(0.349)	(0.300)	(0.587)	(0.506)	(0.413)
$Tr - Con$: lGDP	-3.28	0.06	-3.42	-5.51	-3.36	0.92	-1.75	2.81
$Tr - Con$: lGDPpc	-0.32	1.90	-1.14	-0.38	0.53	0.01	-3.03	0.48
$Tr - Con$: lDiff. GDP	-2.50	-0.29	-3.08	-3.14	0.87	1.35	0.75	3.47
$Tr - Con$: lDiff. GDPpc	-0.62	0.44	-1.15	0.02	0.96	-0.19	-1.41	0.90
N	733356	723300	726606	721847	720271	721277	720923	721385

The weighting undoubtedly improves the comparability of currency unions and the non-currency union control groups substantially. However large differences remain. One important thing to note here is that with such a large sample, it is difficult for these weights to have a meaningful impact on these values in the control group. Comparing the top half of Table 5 with the top half of Table 1 this is quite clear as there is little difference in these sample means. However, the weighted means of treated variables on these observables are indeed much closer to this control set. Notably the EMU gets very close on the GDP measures but is still fairly poor at matching on GDP per capita, where the EMU is much larger than the average in the control set. While the multilateral resistance and dyadic fixed effects in the regression adjustment may help compensate for these remaining differences in the estimation sample, it is important to keep in mind that the first stage does not in general resolve all the problems of inappropriate comparability of controls from Table 1, even if they are substantially improved.

Because of the large sample of controls it is hard, even with quite small weights, to discount the effect of large outliers. This can be the result of extremely low weights (on treated) or high weights (on controls) receiving undue impact due to potentially large weights on a small number of observations or can result from the fact that the large number of control observations that are poor fits are not being properly discounted by the probability weights and are still biasing results. In the next section I explore how these results change when I truncate the sample.

3.4. IPWRA: Truncating the Sample

Extreme outliers are known to impact the IPWRA estimator. While the large sample size should minimize the impact of any one observation, it is also true that the majority of the data is assigned a vanishingly small probability weight for each of the individual estimations of disaggregated currency unions.[12] Indeed this appears to be a problem that increases with the quality of the first stage fit as improvements in the estimates of Table 3 to identify characteristics associated with treatment tend to further discount country-pairs that do not fit the description. Usual rule of thumb estimates consider dropping observations outside of 1%/10% thresholds. Imbens (2004) suggests choosing based on sample size, noting that for a weighted group means estimator that if the maximum allowable weight of any unit is: $1/[N \times (1 - p_{max})]$ for units at the top of the probability spectrum and $1/[N * p_{min}]$ at the bottom. He uses the example that with a sample of 1000, then limiting the impact of any unit to less than 5% then implies a range of [0.02, 0.98]. Thus, because of the large data size I will allow for a much larger range than is used in much of the propensity score matching literature[13] and drop only observations with a probability below 0.001 and above 0.999. This, it turns out, drops an incredible amount of data as the vast majority of observations in the control set are quite poor matches for currency union observations, particularly in the cases of some of the smaller currency unions made up of developing countries. The challenge is that using a smaller range drops still more data, allowing these outliers more influence. Only in the case of the ECCU are observations above this threshold dropped, and there only a small number. While I do not model this trade-off to find an optimal truncation, I will report in this truncated sample the value of the maximum influence, as measured by $1/[N * p_{min}]$. The results for my IPWRA estimates on this truncated sample are reported in Table 6.

---

[12] Fewer are dropped for the aggregate currency union, in part because the treatment model has a hard time fitting the diverse subset of countries and so fewer countries are assigned the near-zero probabilities. [13]Which often looks at smaller observational studies in fields such as medicine.

Table 5: Disaggregated Gravity: IPWRA Truncated Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Currency Union	0.43***							
	(0.02)							
EMU		0.10***						
		(0.02)						
CFA Franc			0.29*					
			(0.15)					
East Caribbean CU				-2.12***				
				(0.50)				
Aussie Dollar					1.00***			
					(0.37)			
British Pound						0.36***		
						(0.05)		
Indian Rupee							1.11*	
							(0.64)	
US Dollar								-0.28
								(0.24)
Exporter-Year FEs	✓	✓	✓	✓	✓	✓	✓	✓
Importer-Year FEs	✓	✓	✓	✓	✓	✓	✓	✓
Dyadic FEs	✓	✓	✓	✓	✓	✓	✓	✓
Max Influence: $1/N^{*}p_{min}$	0.0012	0.0085	0.0079	0.275	0.0743	0.0077	0.165	0.0255
R2	0.9219	0.9688	0.8798	0.9577	0.9924	0.9466	0.9542	0.9693
N	721500	118250	125936	3643	13455	129104	6038	39194

*Standard errors clustered by country-pairs, † p<0.15, * p < 0.10, &** p < 0.05,***_p < 0.01.*

Here the results are extremely different, as one would expect given the radical changes in sample sizes. The overall currency union effect is exactly the same, coming from the fact that almost no data was dropped. As mentioned above, this is likely because the poor fit of the aggregate first stage estimates. Poor fit of treatment can mean very few control observations are "ruled out" by the model and thus only a small fraction fall beneath the 0.001 threshold. Interestingly this is not true for the British Pound CU, despite having a worse pseudo-R^2 than the aggregate union. Looking at the distribution of probabilities this appears to be due to the fact that the estimated logistic model in Table 3 for this union fits low probabilities even for most treated groups, with the mean of probabilities across both treatment and control quite low. In contrast the diversity of currency unions in the aggregate measure makes a poor fit, but finds many comparisons among the controls that, while not exact fits, share many common characteristics, limiting the number of extremely low probability estimates. On the opposite end of the spectrum is the ECCU where the empirical model in Table 3 does a great job identifying treatment units, but has a harder time finding similar comparisons, leading to almost the entire control set falling below this 0.001 threshold for p_{min}.

Looking first at the EMU estimate, for which there is large context in the literature, the parameter is now extremely close to the range of estimates reported in the survey of existing estimates by Polák (2019). Notably my estimates in Table 6 for the EMU are extremely close to those Rose (2017) when limiting the sample to only rich economies. Here I estimate a coefficient of 0.10, while Rose (2017) find a coefficient of 0.11 when estimating the Euro effect among a sample of *"> 75k observations from the rich-country sub-set."* While I did not choose rich countries specifically, the estimation of Equation 4 for selection into the EMU is ultimately trying to match on the bilateral trade relationship of relatively large and relatively rich economies. It would seem my data driven method of selecting a proper control more or less converged on a similar sample to the ad-hoc method used in their work.

Estimates for the ECCU, Aussie Dollar, and Indian Rupee now rely on small sample sizes. Using the arithmetic from Imbens (2004), these now allow for some extreme effects as reported in the table. To test if these probability weights were greatly affecting these estimates, I re-estimated this model on the same sample, while censoring maximum and minimum probabilities to the [0.01,0.99] range. The resulting estimates were not statistically distinguishable, suggesting that the sample reduction choice has driven changes in these estimators, not large weights on outliers. In addition to having such small samples, the estimates for these three have jumped substantially, though the implausibly large estimate for the ECCU is common across all of my estimates, and indeed all of the disaggregated results presented in Glick and Rose (2016).

For three non-EMU disaggregated currency unions the sample reduction exercise does not appear to put undue weight on outliers. For the CFA Franc currency union members this truncation reflects a significant reduction in the currency union effect, but with quite a large 0.29 coefficient, reflecting a 33.6% increase in trade. For the countries using the British pound this effect is only slightly larger than the unweighted sample with a coefficient of 0.36 and implied trade increase of 43.3%. The countries on the US dollar see their small positive coefficient switch to an insignificant negative. In general, for the four disaggregated currency unions where the sample size remains large there is a universal shrinking of trade estimates when moving from the full sample to this truncated one.

Recalling the unweighted and weighted summary statistics in Table 1 and Table 5, one problem with earlier estimates was large sample sizes among non-currency union observations limited the ability of weighted statistics to have meaningful impact on the conditional means of the control sub-populations. This should no longer be the case and so I return once again to the comparison of means across these groups, now including these inverse-probability weights and limiting the samples to those used in Table 6. These are reported in Table 7.

With the exception of the aggregate currency union, where little change in the sample size has taken place, these appear to do a significantly better job of matching these two groups than in the table that included weighting on the full sample. The worst fits are the EMU and the US Dollar. The EMU seems to be a marginal improvement from the differences in these means from Table 5, while still having some substantial differences, while the US dollar is a substantial improvement. Perhaps most interesting in this table is that by truncating the sample the weighted means across these groups are closer to the currency union means in the unweighted summary statistics in Table 1. While the results in Table 5 suggest that weighted full-sample makes the conditional means of the currency union groups look comparable to the non-currency union comparison, here by dropping control units with little comparability the non-currency union groups that remain look much closer to their currency union counterparts. For this reason I generally prefer the estimates on this truncated sample in Table 6.

Table 7: Summary Statistics: Weighted on Truncated Sample

	All CUs	EMU	CFA	ECCU	Aussie Dollar	British Pound	Rupee	US Dollar
Control (CU=0)								
$log(GDP_1 GDP_2)$	47.67	51.53	46.20	42.78	46.22	49.79	49.56	52.45
	(3.860)	(2.440)	(1.499)	(1.470)	(1.088)	(1.322)	(1.506)	(1.833)
$log(GDPpc_1 GDPpc_2)$	16.61	19.98	15.33	18.23	18.44	17.30	14.08	18.55
	(2.880)	(0.722)	(1.306)	(1.012)	(0.612)	(1.076)	(0.996)	(1.551)
$log(\|GDP_1 - GDP_2\|)$	24.18	26.55	23.49	21.69	26.13	27.20	27.28	28.81
	(2.428)	(1.678)	(1.635)	(1.301)	(1.221)	(1.158)	(1.401)	(0.836)
$log(\|GDPpc_1 - GDPpc_2\|)$	7.670	9.395	7.688	8.798	9.823	8.930	6.245	9.297
	(2.016)	(1.230)	(1.585)	(1.267)	(0.779)	(1.401)	(1.391)	(1.309)
Treated (CU=1)								
$log(GDP_1 GDP_2)$	44.39	50.41	45.93	42.32	46.33	49.23	50.48	52.86
	(4.607)	(2.019)	(1.486)	(2.531)	(0.672)	(1.143)	(0.868)	(1.823)
$log(GDPpc_1 GDPpc_2)$	16.29	19.75	15.47	18.39	18.37	17.12	14.39	18.35
	(2.885)	(0.688)	(1.546)	(1.070)	(0.624)	(1.259)	(1.502)	(1.665)
$log(\|GDP_1 - GDP_2\|)$	21.68	26.05	22.63	20.79	26.69	27.14	27.50	29.98
	(2.900)	(1.602)	(0.967)	(1.547)	(0.396)	(0.138)	(1.133)	(0.267)
$log(\|GDPpc_1 - GDPpc_2\|)$	7.051	9.308	7.552	8.454	9.891	8.991	5.302	10.13
	(1.887)	(1.130)	(1.729)	(1.137)	(0.356)	(0.416)	(1.219)	(0.880)
$Tr - Con$: lGDP	-3.28	-1.12	-0.27	-0.46	0.11	-0.56	0.92	0.41
$Tr - Con$: lGDPpc	-0.32	-0.24	0.14	0.17	-0.07	-0.18	0.31	-0.20
$Tr - Con$: lDiff. GDP	-2.50	-0.51	-0.86	-0.90	0.56	-0.06	0.22	1.18
$Tr - Con$: lDiff. GDPpc	-0.62	-0.09	-0.14	-0.34	0.07	0.06	-0.94	0.83
N	722730	118250	125936	3643	13455	129104	6038	39194

3.5. Augmented Inverse Probability Weighting Estimator

The AIPW estimator should provide more efficient estimates than the IPWRA, and while both estimators have the doubly-robust property it is not clear which should perform better under misspecification in both models. These estimates use the probabilities from Table 3 along with the conditional means from the weighted regression estimations in Table 4 to construct the AIPW estimator as defined in Equation 3. This combines a simpler IPW estimator (the first term in Equation 3) with an augmented regression adjustment term, which comes from a weighting of the estimates of the same gravity equation from Equation 1, specified in the same way as for the IPWRA estimator. These are reported in Table 8.

Table 8: AIPW Estimator

	All CUs	EMU	CFA	ECCU	Aussie Dollar	British Pound	Rupee	US Dollar
ATE_{AIPW}	2.45***	0.12	0.22**	-0.81	0.32***	3.88***	0.58***	0.75***
	(0.26)	(0.14)	(0.08)	(0.20)	(0.02)	(0.55)	(0.09)	(0.12)
N	733356	724511	727816	723059	722170	722486	722134	722596

*Standard errors clustered by country-pairs, † p<0.15, * p < 0.10, &** p < 0.05,***_p < 0.01.*

There are two large outliers in these estimates, with positive coefficients that are implausible. These are the aggregate currency union and the British pound. As discussed above these are the two treatments with the worst first-stage fit in Table 3, with the British pound currency union having extremely low probabilities of treatment across even the treated sub-sample. Given that these are estimated on the full sample, one might expect the estimates to be universally closer to those from Table 4. This is not the case. Two disaggregated currency unions, the EMU and CFA, have estimates quite close to their truncated IPWRA estimates in Table 6, which were substantially smaller than those from the full-sample IPWRA. In the case of the EMU this is now imprecisely estimated. The Aussie Dollar and Indian Rupee, on the other hand, have estimates that are extremely close to their non-truncated IPWRA in Table 4, in the case of the Rupee zone almost the same as the estimate from the baseline gravity in Table 2. Notably these are two cases where the limited sample in the truncated case created some cause for concern. The ECCU's persistently large and negative coefficient remains, but is less extreme here.

It is interesting to note that the aggregate currency union effect, British pound, and US dollar were fairly consistent across the full and truncated samples in Table 4 and Table 6, but are dramatically changed in Table 8. This estimator puts substantially more weight on the traditional IPW estimator of weighted group means than the conditional mean estimates from the gravity equations, with these implausible jumps in estimates likely pointing to failure of the modelling of selection into currency unions for these particular disaggregated unions, something that is clearly true for the aggregate and British pound groups. It is not obvious that the fit for the US dollar unions is substantially worse than others, though the balancing of the probabilities in Table 5 and Table 6 for this group remains fairly poor. Because my modelling choices for selection is an atheoretical reduced-form model in Equation 4, I am cautious of reading too closely into results that rely more heavily on the ability of these first stage probabilities to re-weight group means, and while the regression adjustment should stabilize these estimates it clearly struggles to do so in the cases of extremely poor fit in the aggregate CU and British pound. With that said, the EMU, CFA, Australian, and Indian currency unions all have estimates from the AIPW that fit in sensibly with those estimated from the IPWRA estimates.

4. Conclusions

What can be taken away from this exercise? It is not the goal of this paper to advocate for any particular estimates above as the *true* currency union effect on trade, nor even the *true* average treatment effect for a particular disaggregated currency union. Rather I hope to uncover how important the choice and weighting of the comparison group used in such analyses in determining the estimated effects. These results suggest that there should be a change in thinking about how currency union, or any macroeconomic policy estimates are estimated. I have intentionally limited myself to the well studied data of Glick and Rose (2016) as a means of connecting with the large existing literature on these estimates to highlight the potential flaws associated with failure to correct for the endogeneity of selection into a particular policy choice. I see three main takeaways.

First, macroeconomic estimates of trade relationships should focus more on individual policy choices, acknowledging that the choice of Liberia to accept dollars as legal tender is taken under a completely unrelated set of circumstances than the choice of Luxembourg to enter into a multilateral currency regime with Germany. These are likely both different that the decision of Mali and Senegal to band together with neighboring African countries in a multilateral currency union of developing countries with a fixed exchange rate to the Franc/Euro. These differences almost certainly impact the effectiveness that these currency unions have on trade, and comparing them to each other could lead to misleading understanding of the potential gains and losses from currency union formation and destruction.

Another key takeaway is that it is critical to establish a proper counterfactual control group for these policy experiments. The wealth of literature around the EMU impact on trade has made clear that adding huge amounts of data of developing country bilateral trade partners inflates the Euro effect. I have shown above that this largely comes from constructing a sample where the trade partners share few observable characteristics. This is likely equally true for developing country unions, as exported in Campbell and Chentsov (2017), though there is unfortunately much less existing work studying their individual impact to make such connections with.

Finally economists can do better to empirically model the choice of entering into currency arrangements themselves. It would be useful to replicate this cross-country analysis with a more sophisticated dataset on the relevant institutional, demographic, and political factors that might affect first stage selection. This can not only improve the fit of the first-stage probability model over the somewhat crude one used above, but more importantly will better identify those countries who serve as adequate control groups, and likely provide better understanding of *why* a currency union might have better or worse trade impact. Also useful would be to motivate this choice of data with a theoretically consistent model such as that of Alesina and Barro (2002). Even if the majority of data are still poor matches with smaller developing country currency unions, as they were above, improvement of the comparison group will provide more overlap, and therefore should reduce the bias of these estimates. Having a richer dataset to understand the common factors leading to selection into these arrangements will also facilitate a deeper understanding of the trade estimates, by providing better indications of the non-trade related motivations behind the decision to adopt a common currency.

Matching estimators, particularly flexible ones such as the IPWRA and AIPW should be useful in macroeconomic analyses of these estimates going forward. So too might synthetic control estimates that seek to construct appropriate comparison groups from the data when no adequate match is available. This paper shows that even using standard trade models such estimators can go a long way, but the limitations outlined above suggests that more work can be done to improve on these estimates.

References

Aizenman, J. (2016). Optimal currency area: A 20th century idea for the 21st century? Technical report, National Bureau of Economic Research.

Alesina, A. and R. J. Barro (2002). Currency unions. *The Quarterly Journal of Economics 117*(2), 409–436.

Alesina, A., R. J. Barro, and S. Tenreyro (2002). Optimal currency areas. *NBER macroeconomics annual 17*, 301–345.

Baldwin, R. and D. Taglioni (2007). Trade effects of the euro: A comparison of estimators. *Journal of economic integration*, 780–818.

Barro, R. and S. Tenreyro (2007). Economic effects of currency unions. *Economic Inquiry 45*(1), 1–23.

Bayoumi, T. and B. Eichengreen (1993). One money or many? on analyzing the prospects for monetary unification in various parts of the world. Technical report.

Campbell, D. L. (2013). Estimating the impact of currency unions on trade: solving the glick and rose puzzle. *The World Economy 36*(10), 1278–1293.

Campbell, D. L. and A. Chentsov (2017). Breaking badly: the currency union effect on trade.

Chari, V. V., A. Dovis, and P. J. Kehoe (2020). Rethinking optimal currency areas. *Journal of Monetary Economics 111*, 80–94.

Chintrakarn, P. (2008). Estimating the euro effects on trade with propensity score matching. *Review of International Economics 16*(1), 186–198.

Frankel, J. (2010). The estimated trade effects of the euro: why are they below those from historical monetary unions among smaller countries? In *Europe and the Euro*, pp. 169–212. University of Chicago Press.

Glick, R. and A. K. Rose (2002). Does a currency union affect trade? the time-series evidence. *European economic review 46*(6), 1125–1151.

Glick, R. and A. K. Rose (2016). Currency unions and trade: A post-emu reassessment. *European Economic Review 87*, 78–91.

Glick, R. and A. M. Taylor (2010). Collateral damage: Trade disruption and the economic impact of war. *The Review of Economics and Statistics 92*(1), 102–127.

Glynn, A. N. and K. M. Quinn (2010). An introduction to the augmented inverse propensity weighted estimator. *Political analysis 18*(1), 36–56.

Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics*, Volume 4, pp. 131–195. Elsevier.

Head, K., T. Mayer, and J. Ries (2010). The erosion of colonial trade linkages after independence. *Journal of international Economics 81*(1), 1–14.

Hirano, K. and G. W. Imbens (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes research methodology 2*(3), 259–278.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics 86*(1), 4–29.

Jordà, O. and A. M. Taylor (2016). The time for austerity: estimating the average treatment effect of fiscal policy. *The Economic Journal 126*(590), 219–255.

Lunceford, J. K. and M. Davidian (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine 23*(19), 2937–2960.

Millimet, D. L. and R. Tchernis (2009). On the specification of propensity scores, with applications to the analysis of trade policies. *Journal of Business & Economic Statistics 27*(3), 397–415.

Nitsch, V. (2002). Honey, i shrunk the currency union effect on trade. *The World Economy 25*(4), 457–457.

Persson, T. (2001). Currency unions and trade: how large is the treatment effect? *Economic policy 16*(33), 434–448.

Polák, P. (2019). The euro's trade effect: a meta-analysis. *Journal of Economic Surveys 33*(1), 101–124.

Rose, A. (2002). Honey, the currency union effect on trade hasn't blown up. *The World Economy 25*(4), 475–479.

Rose, A. K. (2000). One money, one market: the effect of common currencies on trade. *Economic policy 15*(30), 08–45.

Rose, A. K. (2017). Why do estimates of the emu effect on trade vary so much? *Open Economies Review 28*(1), 1–18.

Rose, A. K. and T. D. Stanley (2005). A meta-analysis of the effect of common currencies on international trade. *Journal of economic surveys 19*(3), 347–365.

Saiki, A. (2005). Asymmetric effect of currency union for developing countries. *Open economies review 16*(3), 227–247.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of econometrics 141*(2), 1281–1301.