LeBow College of Business, Drexel University
School of Economics
Working Paper Series

WP 2017-07

The Currency Union Effect: A PPML Re-assessment with High-Dimensional Fixed Effects

Mario Larch
University of Bayreuth
CESifo & Ifo Institute

Joschka Wanner
University of Bayreuth

Yoto V. Yotov
Drexel University
ERI-BAS & CESifo

Thomas Zylkin
National University of Singapore

Abstract

Recent work on the effects of currency unions (CUs) on trade stresses the importance of using many countries and years in order to obtain reliable estimates. However, for large samples, computational issues limit choice of estimator, leaving an important methodological gap. To address this gap, we unveil an iterative PPML estimator, which flexibly accounts for multilateral resistance, pair-specific heterogeneity, and correlated errors across countries and time. When applied to a comprehensive sample with more than 200 countries trading over 65 years, these innovations flip the conclusions of an otherwise rigorously-specified linear model. Our estimates for both the overall CU effect and the Euro effect specifically are economically small and statistically insignificant. The effect of non-Euro CUs, however, is large and significant. Notably, linear and PPML estimates of the Euro effect increasingly diverge as the sample size grows.

JEL classification: C13; C21; F10; F15, F33

Keywords: Currency Unions; PPML; High-dimensional Fixed Effects

The Currency Union Effect:
A PPML Re-assessment with High-Dimensional Fixed Effects*

Mario Larch[†] Joschka Wanner[‡] Yoto V. Yotov[§] and Thomas Zylkin[¶]

April 29, 2017

Abstract

Recent work on the effects of currency unions (CUs) on trade stresses the importance of using many countries and years in order to obtain reliable estimates. However, for large samples, computational issues limit choice of estimator, leaving an important methodological gap. To address this gap, we unveil an iterative PPML estimator which flexibly accounts for multilateral resistance, pair-specific heterogeneity, and correlated errors across countries and time. When applied to a comprehensive sample with more than 200 countries trading over 65 years, these innovations flip the conclusions of an otherwise rigorously-specified linear model. Our estimates for both the overall CU effect and the Euro effect specifically are economically small and statistically insignificant. The effect of non-Euro CUs, however, is large and significant. Notably, linear and PPML estimates of the Euro effect increasingly diverge as the sample size grows.

JEL Classification Codes: C13, C21, F10; F15, F33
Key Words: Currency Unions; PPML; High-dimensional Fixed Effects

Acknowledgements: Thomas Zylkin is grateful for research support from the NUS Strategic Research Grant (WBS: R-109-000-183-646) awarded to the Global Production Networks Centre (GPN@NUS) for the project titled "Global Production Networks, Global Value Chains, and East Asian Development". Other acknowledgements will be added later.

[†] *Affiliation:* University of Bayreuth, CESifo, and Ifo Institute. *Address:* Universitätsstrasse 30, 95447 Bayreuth, Germany.

[‡] *Affiliation:* University of Bayreuth. *Address:* Universitätsstrasse 30, 95447 Bayreuth, Germany.

[§] *Affiliation:* Drexel University, ERI-BAS, CESifo. *Address:* 1020 G-Hall, 3200 Market Street, Philadelphia, PA, 19104 USA.

[¶] *Affiliation:* National University of Singapore. *Address:* AS2 #06-02, 1 Arts Link, Singapore 117570.

1 Introduction and Motivation

> *To us, a plausible methodology to estimate the currency union effect on trade involves panel estimation with dyadic fixed effects. We [...] await computational advances to be able to estimate the Poisson analogs.(Glick and Rose, 2016, p. 86)*

Writing at the beginning of a transformative period in the empirical study of international trade, Rose (2000) reported the stunning finding that sharing a common currency more than triples trade between countries. While this estimate was regarded as puzzlingly high at the time, it succeeded in stimulating a vibrant and ongoing empirical literature investigating the trade-creating effects of currency unions (CUs), garnering over 2,800 citations (and counting) since its original publication. This literature has notably included frequent re-examinations of the original evidence by Rose himself—such as Glick and Rose (2002, 2016)—as well as fervent interest in whether the European Monetary Union (EMU) in particular, as the largest CU to date, might have had similarly remarkable effects.[1]

Parallel to this literature, the past two decades have seen the development and wide adoption of many new econometric best practices for consistently identifying the determinants of international trade. These have most notably included the use of Poisson Pseudo-maximum Likelihood (PPML) estimation to address issues related to heteroscedasticity and zeroes (Santos Silva and Tenreyro, 2006), time-varying exporter and importer fixed effects to account for changes in multilateral resistance (Feenstra, 2004; Baldwin and Taglioni, 2007), and time-invariant pair fixed effects to absorb unobservable barriers to trade (Baier and Bergstrand, 2007). Aiding these developments, empirical researchers working in trade have also benefited from a new-found consensus on the theoretical underpinnings of the gravity equation (Arkolakis, Costinot, and Rodríguez-Clare, 2012) as well as recent computational

[1] Along with Glick and Rose (2002, 2016), some of Rose's other work in this area includes Rose (2001), Rose (2002), and Rose (2017). Contributions by Persson (2001), Nitsch (2002), Levy-Yeyati (2003), Barro and Tenreyro (2007), and de Sousa (2012) are examples of reactions to Rose's initial finding. Finally, Micco, Stein, and Ordonez (2003), Baldwin and Taglioni (2007), Berger and Nitsch (2008), Santos Silva and Tenreyro (2010a), Eicher and Henn (2011), Olivero and Yotov (2012) and Mika and Zymek (2016) specifically investigate the effect of the EMU. Santos Silva and Tenreyro (2010a) and Rose (2017) survey each of these literatures.

advances which permit swift estimation of linear models with a large number of fixed effects (Carneiro, Guimarães, and Portugal, 2012; Correia, 2016).

Reassuringly, as these new methods have filtered into the literature on currency unions, they have led to more reasonable and reliable estimates. In their latest installment, which emphasizes the use of time-varying exporter and importer fixed effects as well as time-invariant pair fixed effects, Glick and Rose (2016) find—under their most rigorous specification—that CUs generally increase trade by 40%, that CU entry and exit have symmetric effects on trade, and that the EMU—which could not be included in earlier studies—has promoted trade more than other CUs.

Doing their due diligence, Glick and Rose (2016) also experiment with PPML estimation with two-way (exporter-time and importer-time) fixed effects.[2] However, as captured in the opening quote, they are unable to obtain estimates for one particularly important and desirable specification: the case of a PPML model with a *full* set of fixed effects (i.e., also including pair fixed effects).[3] As they rightly point out, two-way estimates without pair fixed effects are generally less reliable, as they do not address the potential endogeneity of CUs.

Thus, in this paper, we pick up where Glick and Rose (2016) leave off. The main technical challenge we overcome is Glick and Rose (2016)'s preference for as large a sample as possible, covering trade between more than 200 countries over 65 years and therefore necessitating the use of more than 50,000 fixed effects. Working with the same data, we describe and implement an iterative PPML algorithm which specifically addresses the computational burden of the three different types of high-dimensional fixed effects ("HDFEs") that need to be

[2]Glick and Rose (2016) include these results in an earlier working paper available online (Glick and Rose, 2015). They still estimate a generally positive "additional effect" for the EMU versus other CUs, but find the overall CU effect disappears over time, echoing an earlier finding by de Sousa (2012).

[3]Even before Glick and Rose (2016), computational challenges with PPML have been quietly simmering for some time. For example, Bratti, De Benedictis, and Santoni (2012) study the impact of immigrants on trade and note that "the very high number of fixed effects causes serious numerical instability to the PPML estimator [...] which failed to reach convergence." Henn and McDonald (2014) find PPML "impracticable [because] convergence of PPML is usually not achieved with fixed effects of a dimensionality as high as ours." And in their services trade handbook, Sauve and Roy (2016) explain that "[u]nfortunately, PPML estimation with several high-dimensional fixed effects led to non-convergence [...] even with the application of different work-around strategies suggested in the recent literature." Dutt, Santacreu, and Traca (2014), Kareem (2014), and Magerman, Studnicka, and Van Hove (2016) share similar frustrations.

computed to obtain unbiased, theory-consistent estimates.[4] In addition, taking advice from Cameron, Gelbach, and Miller (2011) and Egger and Tarlea (2015), we also use standard errors which are clustered on all possible dimensions of the panel—here, exporter, importer, and time—and similarly show how such "multi-way" clustering techniques may be adapted to the HDFE PPML context.

These two methodological changes—changing the underlying estimator and method of clustering—lead to dramatic reversals in what we would otherwise consider the current benchmark estimates from the literature. Unlike the vast majority of studies, we do not find that the average effect of CUs on trade is statistically significant. This is for two main reasons. First, multi-way clustering generally leads to more conservative inferences of all estimates. Using standard, "robust" error corrections, for example, the overall CU effect is positive and measured with high precision. Second, the implications of switching from OLS to PPML are especially pronounced for our estimates of the EMU effect. In fact, for all CUs *other than the EMU,* we find (as much of the literature had until more recently) the effect of sharing a currency has been very large and highly significant, increasing trade by more than 100%. Meanwhile, despite this strong precedent, the effect for the EMU is essentially zero, completely flipping the conclusions of Glick and Rose (2016).

We are not the first researchers to document either a small EMU effect (c.f., Micco, Stein, and Ordonez, 2003; Baldwin and Taglioni, 2007), or, indeed, no EMU effect (c.f., Santos Silva and Tenreyro, 2010a; Olivero and Yotov, 2012). However, other methodological differences aside, these studies have mainly relied on relatively small samples.[5] As Glick and Rose (2016) and Rose (2017) rightly point out, using a sample with many countries

[4]The algorithm we present draws on an earlier method used in Figueiredo, Guimarães, and Woodward (2015) for PPML with two-way HDFEs. It also has several other features not used in this study which other researchers may find useful, such as allowing for pair-specific time trends as well as specifications with added sectoral-level heterogeneity. It was originally programmed by Zylkin (2017) and is available in Stata via ssc: to install, type "`ssc install ppml_panel_sg, replace`".

[5]A notable exception is Mika and Zymek (2016), who also show that the EMU effect vanishes using PPML with many countries. In their paper, the computational issues surrounding PPML are addressed by "artificially balancing" bilateral trade (such that the usual "exporter-time" and "importer-time" FEs become only "country-time" FEs) and by only using more recent years. Glick and Rose (2016) question whether these adjustments lead to truly comparable results. Our findings, however, support those of Mika and Zymek.

and years is in principle always the most sensible approach. However, in practice, it is also this preference that leads to the large difference in estimates. As Santos Silva and Tenreyro (2006) summarize, OLS tends to put relatively more weight on smaller trade flows compared with PPML. Thus, the inclusion of many smaller countries naturally tends to exacerbate the difference between PPML and OLS estimates. We demonstrate this argument by showing how OLS and PPML estimates of the EMU effect continually diverge as increasingly smaller countries are added to the sample. Given the wide-spread use of PPML and the abundance of policy questions that require longitudinal trade data, we hope this simple point will stimulate more rigorous investigations as to why these two leading estimators in the gravity literature give results that can differ so widely.

We now turn to describing our HDFE PPML estimation procedure. The following sections then add our estimates and conclusions.

2 PPML with High-Dimensional Fixed Effects

Let X_{ijt} denote trade flows from exporter i to importer j at time t. $\mathbf{w}_{ijt}$ is a vector containing our covariates of interest, including currency unions and other controls. With exporter-time (λ_{it}), importer-time (ψ_{jt}), and exporter-importer ("pair") fixed effects (μ_{ij}), the estimating equation is

$$X_{ijt} = \exp\left(\lambda_{it} + \psi_{jt} + \mu_{ij} + \mathbf{b}'\mathbf{w}_{ijt}\right) + \nu_{ijt}, \tag{1}$$

where ν_{ijt} denotes the remainder error term. Our goal is to obtain PPML estimates for coefficient vector $\mathbf{b}$ in the presence of these high-dimensional fixed effects. To fix ideas, we first write an expression for the corresponding estimate of $\mathbf{b}$, denoted by $\widehat{\mathbf{b}}$, in the form of a generalized PPML first-order condition:

$$\widehat{\mathbf{b}}: \quad \sum_i \sum_j \sum_t \left[X_{ijt} - \exp\left(\widehat{\lambda}_{it} + \widehat{\psi}_{jt} + \widehat{\mu}_{ij} + \widehat{\mathbf{b}}'\mathbf{w}_{ijt}\right)\right] \mathbf{w}_{ijt} = \mathbf{0}. \tag{2}$$

Noting that the PPML first-order condition for a group fixed effect equates the sum of the dependent variable with the sum of the conditional mean for that group, the remaining first-order conditions associated with (1) may be written as

$$\widehat{\lambda}_{it}: \quad Y_{it} - e^{\widehat{\lambda}_{it}} \sum_j \exp\left(\widehat{\psi}_{jt} + \widehat{\mu}_{ij} + \widehat{\mathbf{b}}'\mathbf{w}_{ijt}\right) = 0, \tag{3}$$

$$\widehat{\psi}_{jt}: \quad X_{jt} - e^{\widehat{\psi}_{jt}} \sum_i \exp\left(\widehat{\lambda}_{it} + \widehat{\mu}_{ij} + \widehat{\mathbf{b}}'\mathbf{w}_{ijt}\right) = 0, \tag{4}$$

$$\widehat{\mu}_{ij}: \quad \sum_t X_{ijt} - e^{\widehat{\mu}_{ij}} \sum_t \exp\left(\widehat{\lambda}_{it} + \widehat{\psi}_{jt} + \widehat{\mathbf{b}}'\mathbf{w}_{ijt}\right) = 0, \tag{5}$$

where $Y_{it} \equiv \sum_j X_{ijt}$ and $X_{jt} \equiv \sum_i X_{ijt}$ respectively denote the sums of all flows associated with each exporter i and importer j at time t.[6]

Along with (2), these equations could be used to solve the complete system in terms of $\widehat{\mathbf{b}}$, $e^{\widehat{\lambda}_{it}}$, $e^{\widehat{\psi}_{jt}}$, and $e^{\widehat{\mu}_{ij}}$ by extending the "zig-zag" algorithm demonstrated in Guimarães and Portugal (2010) for the case of two-way HDFEs. However, to follow more closely the actual methods used, and to emphasize the tight connection linking estimation with theory, it is useful instead to re-write our system of equations in the form of a "structural gravity" model *à la* Anderson and van Wincoop (2003).[7] To do so, first define

$$\Psi_{it} \equiv \frac{Y_{it}/X_{Wt}}{e^{\widehat{\lambda}_{it}}}, \qquad \Phi_{jt} \equiv \frac{X_{jt}/X_{Wt}}{e^{\widehat{\psi}_{jt}}}, \qquad D_{ij} \equiv e^{\widehat{\mu}_{ij}}/X_{Wt}, \tag{6}$$

where $X_{Wt} \equiv \sum_i \sum_j X_{ijt}$ denotes total world trade at time t, to be used as a scaling factor.[8] We make these substitutions because, after plugging these definitions into (1), we arrive at a new version of our estimating equation that closely resembles the famous "structural

[6] Most empirical applications, including the present one, tend not to include "self-trade" (i.e., "X_{ii}") in the estimation. Thus, in our case, Y_{it} is i's total exports and X_{jt} is j's total imports. However, Yotov, Piermartini, Monteiro, and Larch (2016) describe several applications in which including X_{ii} might be appealing. The algorithm allows for either possibility without loss of generality. Furthermore, either approach is compatible with structural gravity (c.f., eq. (17) in French, 2016).

[7] So long as a solution exists, a finer point we discuss further in the Online Appendix, the solution is unique (Gourieroux, Monfort, and Trognon, 1984).

[8] The utility of this scaling factor is that, as in the analogous system used in Anderson and van Wincoop (2003), imposing $D_{ij} = \Phi_{jt} = \Psi_{it} = 1$ (with $\widehat{\mathbf{b}} = 0$) equates to a world where trade frictions do not affect choice of trade partner. This serves as a natural initial guess to use for the solution algorithm.

gravity" equation of Anderson and van Wincoop (2003):

$$X_{ijt} = \left(\frac{Y_{it}X_{jt}}{X_{Wt}}\right)\left(\frac{D_{ij}e^{\widehat{\mathbf{b}}'\mathbf{w}_{ijt}}}{\Psi_{it}\Phi_{jt}}\right) + \varepsilon_{ijt}.$$

And, furthermore, we may also now re-write our system of first-order conditions as follows:

$$\mathbf{0} = \sum_i\sum_j\sum_t\left[X_{ijt} - \left(\frac{Y_{it}X_{jt}}{X_{Wt}}\right)\left(\frac{D_{ij}e^{\widehat{\mathbf{b}}'\mathbf{w}_{ijt}}}{\Psi_{it}\Phi_{jt}}\right)\right]\mathbf{w}_{ijt}, \tag{7a}$$

$$\Psi_{it} = \sum_j \frac{X_{jt}/X_{Wt}}{\Phi_{jt}} D_{ij}e^{\widehat{\mathbf{b}}'\mathbf{w}_{ijt}}, \tag{7b}$$

$$\Phi_{jt} = \sum_i \frac{Y_{it}/X_{Wt}}{\Psi_{it}} D_{ij}e^{\widehat{\mathbf{b}}'\mathbf{w}_{ijt}}, \tag{7c}$$

$$D_{ij} = \frac{\sum_t X_{ijt}}{\sum_t \left(\frac{Y_{it}X_{jt}}{X_{Wt}}\right)\left(\frac{e^{\widehat{\mathbf{b}}'\mathbf{w}_{ijt}}}{\Psi_{it}\Phi_{jt}}\right)}. \tag{7d}$$

In (7b) and (7c), Ψ_{it} and Φ_{jt} are analogs of the "multilateral resistances" from structural gravity. As in Anderson and van Wincoop (2003) (and the vast subsequent literature following Anderson and van Wincoop, 2003), they capture the general equilibrium effects of trade with third countries. The form of these constraints is well-known and Fally (2015) has previously shown they naturally derive from the FOC's of PPML with two-way fixed effects. The new term we add, however, is D_{ij} in (7d), the "pair" fixed effect recommended by Baier and Bergstrand (2007) to address the problem of endogeneity bias due to unobservable heterogeneity across pairs.[9] By the "adding-up" properties of PPML, this last term may be obtained by equating sums of fitted trade flows over time within each pair with sums of actual flows, as in (7d).

With this system in place, the steps to follow are exactly as outlined in (7a)-(7d). That is: (i) given initial guesses for $\{D_{ij}, \Psi_{it}, \Phi_{jt}\}$, compute a solution for $\widehat{\mathbf{b}}$ using (7a); (ii)-(iii) Update Ψ_{it} and Φ_{jt} using (7b) and (7c); (iv) update D_{ij} using (7d); and (v) return to step (i) with new values for $\{D_{ij}, \Psi_{it}, \Phi_{jt}\}$, iterating until convergence.[10]

[9] Specifically, estimates of the effect of CUs will be biased if pairs of countries that "select" into CUs trade more or less than those that do not. We refer readers to Baier and Bergstrand (2007) for further discussion. An added advantage of pair fixed effects is that they control for all observable and unobservable time-invariant trade costs in the structural gravity model.

[10] One way to obtain $\widehat{\mathbf{b}}$ in step (i) would be to solve for it directly via a nonlinear solver. However,

The solution converges rapidly, even for data structures that would ordinarily be too large for direct estimation to be practical. Applying three-way FEs to Glick and Rose (2016)'s data, for example, will require us to account for more than 50,000 fixed effects.[11] Thus, their data will serve as an interesting test, which we now turn to.

3 Re-assessing the Effects of Currency Unions

Following Glick and Rose (2016)'s notation, we define "CU_{ijt}"—a dummy variable equal to 1 if i and j share a common currency in year t—as our main regressor of interest. Thus, we may re-produce Glick and Rose's preferred specification with three-way fixed effects either in its original OLS form,

$$\ln X_{ijt} = \lambda_{it} + \psi_{jt} + \mu_{ij} + \beta \mathbf{z}_{ijt} + \gamma CU_{ijt} + \epsilon_{ijt}, \tag{8}$$

or in the form of our own preferred alternative, using PPML:

$$X_{ijt} = \exp\left(\lambda_{it} + \psi_{jt} + \mu_{ij} + \beta \mathbf{z}_{ijt} + \gamma CU_{ijt}\right) + \nu_{ijt}, \tag{9}$$

where, in either case, $\mathbf{z}_{ijt}$ denotes a set of non-CU controls and the final terms (ϵ_{ijt} and ν_{ijt}) denote residual errors. To motivate our preference for PPML, we note, as Santos Silva and Tenreyro (2006) have, that imposing the OLS moment condition $E[\ln X_{ijt} - \widehat{\ln X}_{ijt}|\cdot] = 0$ does not also imply that $E[X_{ijt} - \widehat{X}_{ijt}|\cdot] = 0$, such that OLS estimates of γ are likely to be biased.[12]

an even more efficient approach is to modify the procedure so that $\widehat{\mathbf{b}}$ can be solved for using iteratively re-weighted least squares (IRLS), as we discuss in the Online Appendix. The Online Appendix also covers other important details such as how to compute clustered standard errors and how to implement the pre-estimation "existence check" recommended by Santos Silva and Tenreyro (2010b).

[11] Table A1 of the Online Appendix summarizes computation times for different sample sizes (both in terms of countries and years considered) for the `ppml`-command of Santos Silva and Tenreyro (2011) and the HDFE `ppml_panel_sg`-command of Zylkin (2017). We also note that the max model size allowed in Stata is only 11,000. For linear specifications this restriction can be circumvented by simple demeaning-transformations not available for non-linear models.

[12] Santos Silva and Tenreyro (2006) provide extensive discussion of this point as well as a comparison study of PPML versus a range of other nonlinear estimators. While PPML implicitly assumes the variance of ν_{ijt} is proportional to the conditional mean, it generally performs adequately even when this assumption is not met. Fernández-Val and Weidner (2016) and Jochmans (forthcoming) have documented favorable

Columns (1) to (3) of Table 1 reproduce the right panel of Table 5 from Glick and Rose (2016). Columns (4) to (6) estimate the same specifications but with PPML. Additionally, we report for each coefficient three types of standard errors: in parentheses we report Huber-White heteroscedasticity-robust standard errors (Huber, 1967; White, 1982) as in Glick and Rose (2016); in brackets we report standard errors clustered by country-pair, the default when using the algorithm of Zylkin (2017); in curly brackets we report multi-way clustered standard errors clustered by exporter, importer, and year. While this turns out to be the most conservative method, we argue it has special salience here, since it actually allows for correlation in the error term within all six possible cluster dimensions $\{i, j, t, it, jt, ij\}$.[13]

Our main observations are as follows. First, note that the main effect for CUs is substantially smaller than in Glick and Rose (2016) (compare, for example, columns (1) and (4).) If standard errors are multi-way clustered, it also becomes statistically insignificant. Given the size of some of the estimates that have been previously in the literature (c.f., Rose, 2000), this is noteworthy.

Second, our PPML estimates for the EMU effect in columns (5) and (6) are even less favorable. When standard errors are clustered either by country-pair or multi-way, the EMU loses significance.[14] Third, however, for all other currency unions *except the EMU*, PPML leads to significant positive effects and the magnitude is nearly tripled versus Glick and Rose (2016), suggesting a trade-promoting effect of $e^{.700} - 1 = 101.3\%$ (versus $e^{.298} - 1 = 34.7\%$). The strong positive result for the "net EMU effect" from Glick and Rose (2016) thus completely reverses, suggesting the EMU has been a major disappointment in this regard.

Finally, other individual CU estimates, shown in column (6), are also affected, to varying

small-sample properties for PPML with two-way FEs. We note that similar investigations for the case of three-way FEs would be valuable additions to the literature.

[13] We have programmed this clustering method into the newest version of the `ppml_panel_sg`-command by way of the `multiway`-option, following the methods described in our Online Appendix. `ppml_panel_sg` also allows user-specified cluster variables for up to four dimensions. See Cameron, Gelbach, and Miller (2011) for a general discussion of multi-way clustering and Egger and Tarlea (2015) for a discussion of this specific point in gravity models

[14] Olivero and Yotov (2012) find the Euro effect is only significant when one accounts for slow, dynamic adjustments over time. On the other hand, Berger and Nitsch (2008) argue the Euro effect is biased upward by not accounting for long-term trends in European trade. For this reason, it is worth mentioning that our results are robust to using pair time-trends, lagged CU and EMU terms, and/or wider time intervals.

degrees. In particular, we see the large PPML estimate for non-EMU CUs is driven by the British £, the French Franc, and "other CUs".[15]

Table 2 then investigates robustness of the findings of Specification (5) with respect to country sample and period of investigation, similar to Glick and Rose (2016)'s Table 8. While no single subsample leads to a positive significant effect for the Euro, the effects for all other currency unions are robust for all samples with a large time span.[16]

For most subsamples, the PPML coefficient estimates for the Euro are substantially smaller than their linear counterparts. However, for the subsample of "Present/future EU" countries we obtain comparable point estimates of -0.305 and -0.267, respectively. This hints at the importance of the number of countries included in the data set, as also emphasized by Rose (2017). Thus, in order to better understand how differences between the PPML and linear estimates evolve with the sample size, in Figure 1, we plot estimates of the linear model and PPML starting with the EU a whole and then adding one country at a time ranked by 2013 GDP, using the full period from 1948-2013. Our estimates reveal that adding more and more (smaller) countries leads to an overall rising OLS estimate, while the PPML estimates stabilize after the inclusion of around 40 additional countries.

The properties of these estimators support a natural interpretation for this divergence. Basically, with the formation of the EMU, trade fell between EMU members relative to trade with both the rest of the EU as well as with the six largest non-EU economies (the US, China, Japan, Brazil, Russia, and India), which together constitute more than two-thirds of world non-EMU GDP. Intra-EMU trade rose, however, relative to trade with smaller partners, starting with the seventh largest non-EU economy (Canada). Intuitively, smaller countries contribute relatively more to the sum of log trade flows (the key moment used to construct

[15]Note we treat missing observations in the Glick and Rose (2016) as missing for both our linear and PPML specifications. As PPML allows zero trade flows, we also run specifications (4)-(6) treating all missing observations as zero trade flows. Table A2 in the Online Appendix reports the results, showing that including zero trade flows hardly affect the estimates. We also perform a Park test (not shown) for all PPML specifications for the validity of a linear model; we receive a p-value of 0.000 (i.e., "reject") in all cases.

[16]Subsamples without observations prior to 1985 include only very few observations of country-pairs leaving or joining non-EMU currency unions. Thus, we find no significant effects (or even cannot identify the effects) for some subsamples. In Table A3 of the Online Appendix we provide the findings for the same subsamples estimated with the linear specification.

fixed effects with OLS) than to the sum of trade flows in levels (the key moment used with PPML). Therefore, OLS places relatively more weight on each additional small country we add to the reference group, even as their contributions to world trade become increasingly marginal.[17]

Forced to choose between these two weighting schemes, we would suggest the effective weights used by PPML are more appropriate. As shown in Fally (2015), and in our own (7b) and (7c), our PPML estimates ensure exact aggregation of the multilateral resistance constraints from a structural gravity model. Since these constraints provide direct theoretical motivation for the form of the estimation, this is highly desirable. Together, PPML's theory-consistent weighting properties and robustness to heteroscedasticity-related concerns provide sound reasons for us to recommend our PPML estimates as a preferred benchmark.

4 Conclusions

We make three main contributions. First, we offer practical methods to overcome important challenges with the estimation of structural gravity models with high-dimensional fixed effects and clustered standard errors using PPML. Second, these innovations lead to very different conclusions about the effects of currency unions on trade, especially for how the Euro has failed to promote trade while non-Euro CUs emphatically have.

Third, we identify a cautionary example where OLS and PPML gravity estimates differ to an especially dramatic degree. We relate this difference to the number of small countries included in our sample and the effective weights used by each estimator. While we have several reasons for preferring our PPML estimates, we find it notable that computational issues limit choice of estimator precisely when this choice seems to matter most. Thus, we, too, await future advances, which we hope will provide more rigorous guidance for estimating three-way gravity models.

[17] See Fally (2015) for a discussion of the moments used to construct fixed effects under both OLS and PPML.

Table 1: Linear Specification vs. PPML

	Linear Specifications			PPML		
	All CUs	Disagg. EMU	Disagg. CUs	All CUs	Disagg. EMU	Disagg. CUs
	(1)	(2)	(3)	(4)	(5)	(6)
All CUs	0.343			0.130		
	(0.018)***			(0.010)***		
	[0.049]***			[0.042]***		
	{0.080}***			{0.081}		
EMU		0.429	0.432		0.030	0.027
		(0.021)***	(0.021)***		(0.010)***	(0.010)***
		[0.056]***	[0.056]***		[0.042]	[0.041]
		{0.149}***	{0.149}***		{0.092}	{0.091}
All Non-EMU CUs		0.298			0.700	
		(0.025)***			(0.025)***	
		[0.069]***			[0.107]***	
		{0.097}***			{0.172}***	
CFA Franc Zone			0.583			0.137
			(0.100)***			(0.108)
			[0.212]***			[0.339]
			{0.186}***			{0.307}
East Caribbean CU			-1.637			-1.014
			(0.106)***			(0.081)***
			[0.255]***			[0.278]***
			{0.334}***			{0.319}***
Aussie $			0.389			0.168
			(0.196)**			(0.121)
			[0.325]			[0.293]
			{0.248}			{0.282}
British £			0.554			1.004
			(0.034)***			(0.034)***
			[0.086]***			[0.143]***
			{0.101}***			{0.234}***
French Franc			0.874			2.096
			(0.083)***			(0.062)***
			[0.245]***			[0.219]***
			{0.269}***			{0.302}***
Indian Rupee			0.522			0.082
			(0.115)***			(0.149)
			[0.318]			[0.373]
			{0.110}***			{0.308}
US $			-0.051			0.014
			(0.063)			(0.022)
			[0.164]			[0.068]
			{0.229}			{0.065}
Other CUs			-0.104			0.788
			(0.058)*			(0.052)***
			[0.179]			[0.187]***
			{0.247}			{0.247}***
RTAs	0.395	0.392	0.389	0.167	0.169	0.168
	(0.009)***	(0.010)***	(0.010)***	(0.009)***	(0.009)***	(0.009)***
	[0.024]***	[0.024]***	[0.024]***	[0.040]***	[0.039]***	[0.039]***
	{0.062}***	{0.061}***	{0.061}***	{0.076}**	{0.075}**	{0.075}**
CurCol	0.262	0.275	0.248	0.733	0.545	0.303
	(0.032)***	(0.032)***	(0.033)***	(0.059)***	(0.050)***	(0.042)***
	[0.108]**	[0.109]**	[0.112]**	[0.270]***	[0.220]**	[0.164]*
	{0.155}*	{0.159}*	{0.170}	{0.288}**	{0.251}**	{0.148}**
N	877736	877736	877736	879794	879794	879794

Notes: Columns (1) to (3) of this table reproduce the right panel of Table 5 from Glick and Rose (2016). Columns (4) to (6) estimate the same specifications but with PPML. 879,794 observations for more than 200 countries for the years 1948 to 2013. All columns include (roughly) 11,000 exporter-time, 11,000 importer-time, and 32,000 pair FEs. Robust standard errors in parentheses. Standard errors clustered by country pairs in brackets. Standard errors clustered by exporter, importer, and year in curly brackets. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

Table 2: PPML Estimation of Different Subsamples

	1948-2013	1985-2013	1995-2013	1948-2005	1985-2005	1995-2005
				All countries		
EMU	0.030	0.006	0.010	-0.055	-0.063	-0.052
	(0.010)***	(0.010)	(0.013)	(0.013)***	(0.010)***	(0.011)***
	[0.042]	[0.032]	[0.027]	[0.040]	[0.029]**	[0.022]**
	{0.092}	{0.058}	{0.038}	{0.082}	{0.050}	{0.034}
All Non-EMU CUs	0.700	0.084	0.052	0.685	-0.002	0.009
	(0.025)***	(0.027)***	(0.031)*	(0.025)***	(0.030)	(0.035)
	[0.107]***	[0.078]	[0.088]	[0.103]***	[0.067]	[0.074]
	{0.172}***	{0.073}	{0.087}	{0.156}***	{0.056}	{0.072}
				Industrial countries plus present/future EU		
EMU	-0.138	-0.055	-0.009	-0.200	-0.122	-0.075
	(0.012)***	(0.011)***	(0.014)	(0.017)***	(0.012)***	(0.012)***
	[0.047]***	[0.036]	[0.029]	[0.049]***	[0.035]***	[0.026]***
	{0.081}*	{0.055}	{0.037}	{0.080}**	{0.046}***	{0.034}**
All Non-EMU CUs	1.159	-0.188	0.007	1.066	-0.050	0.018
	(0.043)***	(0.147)	(0.268)	(0.041)***	(0.145)	(0.172)
	[0.181]***	[0.304]	[0.292]	[0.166]***	[0.290]	[0.212]
	{0.270}***	{0.366}	{0.160}	{0.232}***	{0.282}	{0.095}
				Upper income (GDP p/c $\geq$ \$ 12,736)		
EMU	-0.076	-0.027	-0.002	-0.134	-0.089	-0.063
	(0.012)***	(0.011)**	(0.015)	(0.015)***	(0.012)***	(0.013)***
	[0.043]*	[0.036]	[0.030]	[0.043]***	[0.035]**	[0.027]**
	{0.073}	{0.052}	{0.037}	{0.066}**	{0.043}**	{0.032}**
All Non-EMU CUs	0.762			0.743		
	(0.130)***			(0.107)***		
	[0.186]***			[0.152]***		
	{0.232}***			{0.194}***		
				Rich Big (GDP$\geq$\$ 10bn, GDP p/c$\geq$\$ 10k)		
EMU	-0.055	-0.025	-0.004	-0.108	-0.088	-0.073
	(0.012)***	(0.011)**	(0.015)	(0.015)***	(0.012)***	(0.013)***
	[0.044]	[0.036]	[0.030]	[0.044]**	[0.034]***	[0.027]***
	{0.077}	{0.053}	{0.038}	{0.073}	{0.043}**	{0.032}**
All Non-EMU CUs	1.312			1.223		
	(0.059)***			(0.055)***		
	[0.184]***			[0.158]***		
	{0.321}***			{0.256}***		
				OECD		
EMU	-0.151	-0.059	-0.017	-0.203	-0.117	-0.067
	(0.013)***	(0.013)***	(0.018)	(0.016)***	(0.013)***	(0.014)***
	[0.046]***	[0.042]	[0.036]	[0.045]***	[0.037]***	[0.029]**
	{0.072}**	{0.068}	{0.048}	{0.055}***	{0.046}**	{0.038}*
All Non-EMU CUs	1.227			1.180		
	(0.061)***			(0.059)***		
	[0.220]***			[0.187]***		
	{0.434}***			{0.365}***		
				Present/future EU		
EMU	-0.305	-0.068	0.021	-0.448	-0.192	-0.060
	(0.017)***	(0.014)***	(0.017)	(0.026)***	(0.018)***	(0.017)***
	[0.064]***	[0.036]*	[0.033]	[0.072]***	[0.045]***	[0.037]
	{0.099}***	{0.055}	{0.041}	{0.124}***	{0.084}**	{0.063}
All Non-EMU CUs	1.157			1.131		
	(0.054)***			(0.052)***		
	[0.233]***			[0.206]***		
	{0.517}**			{0.469}**		

Notes: This table reports robustness estimates of the findings of Specification (5) with respect to country sample and period of investigation, as in Table 8 of Glick and Rose (2016). Robust standard errors in parentheses. Standard errors clustered by country pairs in brackets. Standard errors clustered by exporter, importer, and year in curly brackets. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

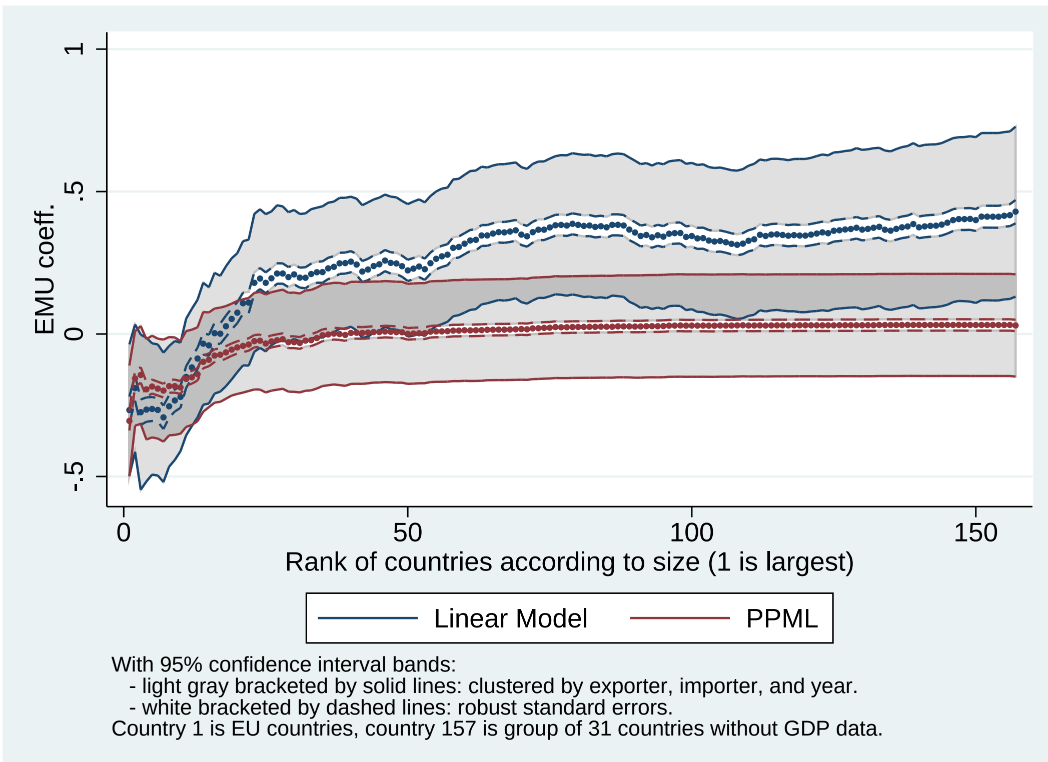

Figure 1: The Effect of the Country Sample on EMU Coefficient Estimates

References

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," in *Center for Economic Studies, US Census Bureau*, vol. 2002-06.

Anderson, J. E., and E. van Wincoop (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93(1), 170–192.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012): "New Trade Models, Same Old Gains?," *American Economic Review*, 102(1), 94–130.

Baier, S. L., and J. H. Bergstrand (2007): "Do Free Trade Agreements Actually Increase Members' International Trade?," *Journal of International Economics*, 71(1), 72–95.

Baldwin, R., and D. Taglioni (2007): "Trade Effects of the Euro: A Comparison of Estimators," *Journal of Economic Integration*, 22(4), 780–818.

Barro, R., and S. Tenreyro (2007): "Economic Effects of Currency Unions," *Economic Inquiry*, 45, 1–23.

Berger, H., and V. Nitsch (2008): "Zooming Out: The Trade Effect of the Euro in Historical Perspective," *Journal of International Money and Finance*, 27(8), 1244–1260.

Bratti, M., L. De Benedictis, and G. Santoni (2012): "On the Pro-Trade Effects of Immigrants," IZA Discussion Papers 6628, Institute for the Study of Labor (IZA).

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference with Multiway Clustering," *Journal of Business & Economic Statistics*, 29(2), 238–249.

Carneiro, A., P. Guimarães, and P. Portugal (2012): "Real Wages and the Business Cycle: Accounting for Worker, Firm, and Job Title Heterogeneity," *American Economic Journal: Macroeconomics*, 4(2), 133–152.

Correia, S. (2016): "REGHDFE: Stata Module to Perform Linear or Instrumental-Variable Regression Absorbing Any Number of High-Dimensional Fixed Effects," https://ideas.repec.org/c/boc/bocode/s457874.html.

de Sousa, J. (2012): "The Currency Union Effect on Trade is Decreasing Over Time," *Economics Letters*, 117(3), 917–920.

Dutt, P., A. M. Santacreu, and D. Traca (2014): "The Gravity of Experience," Working Papers 2014-41, Federal Reserve Bank of St. Louis.

Egger, P. H., and F. Tarlea (2015): "Multi-Way Clustering Estimation of Standard Errors in Gravity Models," *Economics Letters*, 134, 144–147.

Eicher, T., and C. Henn (2011): "One Money, One Market: A Revised Benchmark," *Review of International Economics*, 19(3), 419–435.

Fally, T. (2015): "Structural Gravity and Fixed Effects," *Journal of International Economics*, 97(1), 76–85.

Feenstra, R. C. (2004): *Advanced International Trade: Theory and Evidence*. Princeton University Press, Princeton, NJ.

Fernández-Val, I., and M. Weidner (2016): "Individual and Time Effects in Nonlinear Panel Models with Large N, T," *Journal of Econometrics*, 192(1), 291–312.

Figueiredo, O., P. Guimarães, and D. Woodward (2015): "Industry Localization, Distance Decay, and Knowledge Spillovers: Following the Patent Paper Trail," *Journal of Urban Economics*, 89, 21–31.

French, S. (2016): "The Composition of Trade Flows and the Aggregate Effects of Trade Barriers," *Journal of International Economics*, 98, 114–137.

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 (2015): "Currency Unions and Trade: A Post-EMU Mea Culpa," Discussion paper, National Bureau of Economic Research.

Glick, R., and A. K. Rose (2016): "Currency Unions and Trade: A Post-EMU Reassessment," *European Economic Review*, 87, 78–91.

Gourieroux, A. C., A. Monfort, and A. Trognon (1984): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52(3), 681–700.

Guimarães, P. (2016): "POI2HDFE: Stata Module to Estimate a Poisson Regression with Two High-Dimensional Fixed Effects," https://ideas.repec.org/c/boc/bocode/s457777.html.

Guimarães, P., and P. Portugal (2010): "A Simple Feasible Procedure to Fit Models with High-Dimensional Fixed Effects," *Stata Journal*, 10(4), 628–649.

Henn, C., and B. McDonald (2014): "Crisis Protectionism: The Observed Trade Impact," *IMF Economic Review*, 62(1), 77–118.

Huber, P. J. (1967): "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium*, ed. by J. Neyman, pp. 221–233. University of California Press, Berkeley, CA.

Jochmans, K. (forthcoming): "Two-Way Models for Gravity," *Review of Economics and Statistics*.

Kareem, F. O. (2014): "Modeling and Estimation of Gravity Equation in the Presence of Zero Trade: A Validation of Hypotheses Using Africa's Trade Data," Discussion paper.

Levy-Yeyati, E. (2003): "On the Impact of a Common Currency on Bilateral Trade," *Economics Letters*, 79(1), 125–129.

Magerman, G., Z. Studnicka, and J. Van Hove (2016): "Distance and Border Effects in International Trade: A Comparison of Estimation Methods," *Economics - The Open-Access, Open-Assessment E-Journal*, 10, 1–31.

Micco, A., E. Stein, and G. Ordonez (2003): "The Currency Union Effect on Trade: Early Evidence from EMU," *Economic Policy*, 18(37), 315–356.

Mika, A., and R. Zymek (2016): “Friends Without Benefits? New EMU Members and the “Euro Effect" on Trade,” in *Edinburgh School of Economics*, vol. 269.

Nelder, J. A., and R. W. M. Wedderburn (1972): “Generalized Linear Models,” *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370.

Nitsch, V. (2002): “Honey, I Shrunk the Currency Union Effect on Trade,” *The World Economy*, 25(4), 457–474.

Olivero, M. P., and Y. V. Yotov (2012): “Dynamic Gravity: Endogenous Country Size and Asset Accumulation,” *Canadian Journal of Economics*, 45(1), 64–92.

Persson, T. (2001): “Currency Unions and Trade: How Large is the Treatment Effect?,” *Economic Policy*, 16(33), 433–448.

Rose, A. K. (2000): “One Money One Market: Estimating the Effects of Common Currencies on Trade,” *Economic Policy*, 15(1), 9–48.

——— (2001): “Currency Unions and Trade: The Effect is Large,” *Economic Policy*, 16(33), 449–461.

——— (2002): “Honey, the Currency Union Effect on Trade Hasn’t Blown Up,” *The World Economy*, 25(4), 475–479.

——— (2017): “Why do Estimates of the EMU Effect on Trade Vary so Much?,” *Open Economies Review*, 28(1), 1–18.

Santos Silva, J. M. C., and S. Tenreyro (2006): “The Log of Gravity,” *Review of Economics and Statistics*, 88(4), 641–658.

——— (2010a): “Currency Unions in Prospect and Retrospect,” *Annual Review of Economics*, 2(1), 51–74.

——— (2010b): “On the Existence of the Maximum Likelihood Estimates in Poisson Regression,” *Economics Letters*, 107(2), 310–312.

——— (2011): “poisson: Some Convergence Issues,” *Stata Journal*, 11(2), 207–212.

Sauve, P., and M. Roy (2016): *Research Handbook on Trade in Services*. Edward Elgar Publishing Ltd., UK.

White, H. (1982): “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 50(1), 1–25.

Yotov, Y. V., R. Piermartini, J.-A. Monteiro, and M. Larch (2016): *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*. World Trade Organization, Geneva.

Zylkin, T. (2017): “PPML_PANEL_SG: Stata Module to Estimate “Structural Gravity” Models via Poisson PML,” https://ideas.repec.org/c/boc/bocode/s458249.html.

Online Appendix for "The Currency Union Effect: A PPML Re-assessment with High-Dimensional Fixed Effects"

by Mario Larch, Joschka Wanner, Yoto V. Yotov, Thomas Zylkin

This Online Appendix elaborates on several important considerations such as how to obtain multi-way clustered standard errors and how to verify before estimation that valid estimates do indeed exist. It is in part intended to serve as additional technical documentation for interested readers seeking to work with or extend the machinery used in `ppml_panel_sg`. All procedures described here can be verified to reproduce results produced by other widely-used routines. See the supporting material included with Zylkin (2017) for examples. Further, we provide some additional results and robustness checks.

Iteratively re-weighted least squares algorithm. The IRLS version of the algorithm is analogous to typical IRLS estimation in that it repeatedly utilizes weighted least squares estimation (of a particular form specific to the estimator being used), which is continuously updated as new estimates are produced, until both weights and estimates eventually converge. An IRLS approach is thus easily embedded within the broad approach described in the paper.

For IRLS estimation of a PPML model, it is necessary to first define an adjusted dependent variable—call it $\widetilde{X}_{ijt}$—which is given by:

$$\widetilde{X}_{ijt} = \frac{X_{ijt} - \widehat{X}_{ijt}}{\widehat{X}_{ijt}} + \widehat{\mathbf{b}}'\mathbf{w}_{ijt},$$

For PPML, the relevant weighting matrix for the estimation is simply given by the conditional mean $\widehat{X}_{ijt}$. Thus, given $\widehat{X}_{ijt}$ and $\widetilde{X}_{ijt}$, an updated value for $\widehat{\mathbf{b}}$ can be simply computed as:

$$\widehat{\mathbf{b}} = \left[\mathbf{W}'\widehat{\mathbf{X}}\mathbf{W}\right]^{-1}\mathbf{W}\widehat{\mathbf{X}}\widetilde{\mathbf{X}},$$

where $\widehat{\mathbf{X}}$ is a diagonal weighting matrix with elements $\widehat{X}_{ijt}$ on its main diagonal and $\mathbf{W}$ is the matrix of main covariates $\mathbf{w}_{ijt}$. As in a more-typical IRLS loop, the weighting matrix is updated repeatedly as each new iteration of $\widehat{\mathbf{b}}$ implies a new conditional mean.[18] What must be added here are the intermediate steps needed to compute Ψ, Φ, and D, which follow from (7b)-(7d). Iterating repeatedly on these objects, along with $\widehat{\mathbf{b}}$, will eventually converge to the correct conditional mean, weighting matrix, and PPML estimates for $\widehat{\mathbf{b}}$. Since the algorithm requires repeated iteration anyway, the IRLS method is always the most efficient approach versus solving the first-order condition for $\widehat{\mathbf{b}}$ exactly each time through the loop.[19]

[18]For clarity, $\widetilde{X}_{ijt}$ is derived from a first-order Taylor approximation of the PPML FOC for $\widehat{\mathbf{b}}$ around $\widehat{\mathbf{b}}^0$, where $\widehat{\mathbf{b}}^0$ denotes the current guess for $\widehat{\mathbf{b}}$. The use of $\widehat{X}$ as a weighting matrix also follows from this approximation. For a reference, see Nelder and Wedderburn (1972).

[19]The adoption of IRLS in `ppml_panel_sg` was inspired by the use of a similar principle—albeit in an altogether very different procedure—in the latest version of `poi2hdfe`, by Guimarães (2016).

Three-way within transformation. A useful prior for the rest of these notes is the notion of a three-way "within-transformation", generalizing the two-way procedures of Abowd, Creecy, and Kramarz (2002) and Guimarães and Portugal (2010) and as may be applied via the `hdfe` algorithm of Correia (2016).

Let each of the "main" (non-fixed effect) regressors of the vector $\mathbf{w}_{ijt}$ on the right hand side be denoted by w_{ijt}^k, with superscript k indexing the kth regressor. The idea is to (iteratively) regress each w_{ijt}^k on the complete set of fixed effects. Doing so results in a new set of "partialed-out" (or "within-transformed") versions of w_{ijt}^k, which have been removed of any partial correlation with the set of fixed effects. For the current three-way HDFE context—with it, jt, and ij fixed effects—the needed within-transformation for each w_{ijt}^k is given by the following system of equations:

$$\sum_j \left(w_{ijt}^k - \widetilde{\lambda}_{it}^k - \widetilde{\psi}_{jt}^k - \widetilde{\mu}_{ij}^k \right) = 0 \quad \forall i, t \tag{A1a}$$

$$\sum_i \left(w_{ijt}^k - \widetilde{\lambda}_{it}^k - \widetilde{\psi}_{jt}^k - \widetilde{\mu}_{ij}^k \right) = 0 \quad \forall j, t \tag{A1b}$$

$$\sum_t \left(w_{ijt}^k - \widetilde{\lambda}_{it}^k - \widetilde{\psi}_{jt}^k - \widetilde{\mu}_{ij}^k \right) = 0 \quad \forall i, j, \tag{A1c}$$

where (A1a)-(A1c) are derived from the first-order conditions from an OLS regression of w^k on a set of fixed effects $\{\widetilde{\lambda}_{it}^k, \widetilde{\psi}_{it}^k, \widetilde{\mu}_{ij}^k\}$. Either by using "zig-zag" iteration methods or via the more sophisticated algorithm of Correia (2016), this system is easily solved even for a large number of fixed effects. The resulting, now-transformed regressors, which we will denote as $\widetilde{w}^k$, are given by:

$$\widetilde{w}_{ijt}^k = w_{ijt}^k - \widetilde{\lambda}_{it}^k - \widetilde{\psi}_{jt}^k - \widetilde{\mu}_{ij}^k.$$

Variations of this within-transformation procedure will come into play in the discussion that follows of how we construct standard errors as well as how we implement the "check for existence" recommended by Santos Silva and Tenreyro (2010b). Thus, these basic mechanics will be helpful to keep in mind.

Standard errors. The construction of standard errors largely follows the exposition in the Appendix of Figueiredo, Guimarães, and Woodward (2015), which we extend to the case of three-way HDFEs with multi-way clustering. Let $\sum_{i,j,t}$ denote a sum over all observations and let $\mathbf{x}_{ijt}$ denote the vector of all covariates associated with observation ijt, including all 0/1 dummy variables associated with each fixed effect. The estimated "robust" variance-covariance (VCV) matrix for our PPML estimates that we need to construct is given by

$$\widehat{\mathbf{V}}_{rob} = \underbrace{\left[\sum_{i,j,t} \widehat{X}_{ijt} \mathbf{x}_{ijt} \mathbf{x}_{ijt}' \right]^{-1}}_{\widehat{\mathbf{V}}} \times \underbrace{\left[\sum_{i,j,t} \left(X_{ijt} - \widehat{X}_{ijt} \right)^2 \mathbf{x}_{ijt} \mathbf{x}_{ijt}' \right]}_{\mathbf{M}} \times \underbrace{\left[\sum_{i,j,t} \widehat{X}_{ijt} \mathbf{x}_{ijt} \mathbf{x}_{ijt}' \right]^{-1}}_{\widehat{\mathbf{V}}}, \tag{A2}$$

where $\widehat{\mathbf{V}}$ is proportional to the usual (uncorrected) Poisson MLE VCV matrix and $\widehat{X}_{ijt}$ is the conditional mean from our regression. The middle term, $\mathbf{M}$, provides a heteroscedasticity correction.

While we can compute the matrix $\sum_{i,j,t} \widehat{X}_{ijt}\mathbf{x}_{ijt}\mathbf{x}'_{ijt}$, inversion of this matrix is potentially infeasible due to the large dimension of $\mathbf{x}_{ijt}$. The problem is simplified, however, by recognizing we are only interested in the submatrix of $\widehat{\mathbf{V}}$ that pertains to $\widehat{\mathbf{b}}$, the coefficients for our non-fixed effect regressors. Call this submatrix $\widehat{\mathbf{V}}^*$. To obtain $\widehat{\mathbf{V}}^*$, we make use of the following two "tricks": (i) the $\widehat{\mathbf{V}}$ that appears in (A2) is proportional to the VCV matrix that would be produced by *any* weighted least squares regression using $\mathbf{x}_{ijt}$ as covariates and $\sqrt{\widehat{X}_{ijt}}$ as weights; (ii) By the Frisch-Waugh-Lovell theorem, the dimensionality of an HDFE linear regression can be easily reduced by first applying a within-transformation (a weighted one in this case).

We thus proceed in two steps. First, using a weighted version of our within-transformation procedure, we regress each weighted regressor $\sqrt{\widehat{X}_{ijt}}w^k_{ijt}$ on a set of exporter-time, importer-time, and exporter-importer fixed effects, which themselves must also be weighted by $\sqrt{\widehat{X}_{ijt}}$. The system of equations associated with this operation may be written as

$$\sum_j \widehat{X}_{ijt}\left(w^k_{ijt} - \widetilde{\lambda}^{k*}_{it} - \widetilde{\psi}^{k*}_{jt} - \widetilde{\mu}^{k*}_{ij}\right) = 0 \quad \forall i, t \tag{A3a}$$

$$\sum_i \widehat{X}_{ijt}\left(w^k_{ijt} - \widetilde{\lambda}^{k*}_{it} - \widetilde{\psi}^{k*}_{jt} - \widetilde{\mu}^{k*}_{ij}\right) = 0 \quad \forall j, t \tag{A3b}$$

$$\sum_t \widehat{X}_{ijt}\left(w^k_{ijt} - \widetilde{\lambda}^{k*}_{it} - \widetilde{\psi}^{k*}_{jt} - \widetilde{\mu}^{k*}_{ij}\right) = 0 \quad \forall i, j, \tag{A3c}$$

where $\{\widetilde{\lambda}^{k*}_{it}, \widetilde{\psi}^{k*}_{it}, \widetilde{\mu}^{k*}_{ij}\}$ are the fixed effects terms we now need to solve for. Despite the presence of $\widehat{X}_{ijt}$ in (A3a)-(A3c), the basic principles and methods to solve are no different than with (A1a)-(A1c).

The transformed regressors we need for our auxiliary regression—call these $\widetilde{w}^{k*}_i$—are given by

$$\widetilde{w}^{k*}_{ijt} = \sqrt{\widehat{X}_{ijt}}\left(w^k_{ijt} - \widetilde{\lambda}^{k*}_{it} - \widetilde{\psi}^{k*}_{jt} - \widetilde{\mu}^{k*}_{ij}\right).$$

With these residuals in hand, the second step is to now perform the following OLS regression:

$$X_{ijt} = \sum_k a_k \widetilde{w}^{k*}_{ijt} + u_i \tag{A4}$$

The estimates obtained from this regression are irrelevant. The main point is that, after employing the two "tricks" mentioned above, the VCV matrix from (A4) will be equal to $s^2 \times \widehat{\mathbf{V}}^*$, where s^2 is the usual mean squared error from the linear regression.

Finally, now that we have $\widehat{\mathbf{V}}^*$, the full, heteroscedasticity-robust VCV matrix for our main regressors can be computed as

$$\widehat{\mathbf{V}}^*_{rob} = \widehat{\mathbf{V}}^* \times \mathbf{M}^* \times \widehat{\mathbf{V}}^*,$$

where the middle term,

$$\mathbf{M}^* = \left[\sum_{i,j,t} \frac{\left(X_{ijt} - \widehat{X}_{ijt}\right)^2}{X_{ijt}} \widetilde{\mathbf{w}}^*_{ijt} \widetilde{\mathbf{w}}^{*\prime}_{ijt} \right],$$

must be adjusted to take into account the fact that each $\widetilde{w}^{k*}_{ijt}$ is weighted by $\sqrt{\widehat{X}_{ijt}}$.

Multi-way clustering. The multi-way clustered VCV matrix takes the form

$$\widehat{\mathbf{V}}^*_{clus} = \widehat{\mathbf{V}}^* \mathbf{M}^*_{clus} \widehat{\mathbf{V}}^*,$$

where $\widehat{\mathbf{V}}^*$ is calculated in the exact same way as described above. For the matrix $\mathbf{M}^*_{clus}$, we follow Cameron, Gelbach, and Miller (2011), taking into account that we are still dealing only with a submatrix of the overall matrix $\widehat{\mathbf{V}}$, and calculate it as follows:

$$\mathbf{M}^*_{clus} = \sum_{||\mathbf{r}||=k, \mathbf{r} \in R} (-1)^{k+1} \tilde{\mathbf{M}}^*_{\mathbf{r}},$$

with

$$\tilde{\mathbf{M}}^*_{\mathbf{r}} = \sum_l \sum_m \frac{\left(X_l - \widehat{X}_l\right)}{\sqrt{X_l}} \frac{\left(X_m - \widehat{X}_m\right)}{\sqrt{X_m}} \widetilde{\mathbf{w}}^*_l \widetilde{\mathbf{w}}^{*\prime}_m I_{\mathbf{r}}(l, m) \qquad \mathbf{r} \in R,$$

where the set $R \equiv \{\mathbf{r} : r_d \in \{0, 1\}, d = 1, 2, ..., D, \mathbf{r} \neq \mathbf{0}\}$, where D is the number of dimensions of clustering and the elements of R index whether two observations are joint members of at least one cluster. l and m denote specific ijt-observations. $I_{\mathbf{r}}(l, m)$ takes the value one if observations l and m are both members of all clusters for which $r_d = 1$. $||\mathbf{r}||$ denotes the ℓ_1-norm of the vector $\mathbf{r}$.

Check for existence. As illuminated in Santos Silva and Tenreyro (2010b), depending on the configuration of the data, estimates from Poisson regressions may not actually exist. Specifically, if two or more regressors are perfectly collinear over the subsample where the dependent variable is non-zero, researchers are advised to carefully investigate each "implicated" regressor to see if it can be included in their model. Otherwise, estimation routines may result in spurious estimates, or even no estimates at all.[20]

With multiple high-dimensional fixed effects, implementing the checks favored by Santos Silva and Tenreyro (2010b) may seem a daunting task, since collinearity checks across all the different fixed effects to determine whether one or more are "implicated" may be computationally expensive and/or conceptually difficult, especially when there are more than two HDFEs. In addition, it is also necessary to check whether each individual regressor is collinear over $X_{ijt} > 0$ with the complete set of fixed effects, as well as whether any subset of fixed effect and non-fixed effect regressors are collinear over $X_{ijt} > 0$.

[20] Note this is a different issue altogether than the standard issue of "perfect collinearity" and can be significantly more difficult to detect. See Santos Silva and Tenreyro (2010b) for a simple example of a model with non-collinear regressors which does not have a solution.

Fortunately, however, it turns out these issues are quickly and easily resolved by (i) applying the within-transformation technique described above and (ii) recognizing that fixed effects themselves only present an issue under easily-identifiable circumstances. To see this, let "$\widetilde{w}^k_{ijt|X>0}$" denote the within-transformed version of each non-fixed effect regressor w_k after performing a within-transformation (only this time restricted to the subsample $X_{ijt} > 0$). After applying the within-transformation, these $\widetilde{w}^k_{ijt|X>0}$'s now only contain the residual variation in each w^k_{ijt} over $X_{ijt} > 0$ that is uncorrelated with the set of fixed effects. Thus, any individual $\widetilde{w}^k_{ijt|X>0}$ that is uniformly zero should be considered "implicated", since this only occurs if w^k_{ijt} is perfectly collinear with the set of fixed effects over $X_{ijt} > 0$. Furthermore, it is now a simple matter to apply a standard collinearity check among the remaining $\widetilde{w}^k_{ijt|X>0}$ to test for joint collinearity over $X_{ijt} > 0$, taking into account all possible correlations with the set of fixed effects.

That still leaves the matter of collinearity among the potentially very many fixed effects, which may seem the most difficult step of all. However, Santos Silva and Tenreyro (2010b) also clarify that it should always be possible to include any regressor that has "reasonable overlap" in the values that it takes over both the $X_{ijt} > 0$ and $X_{ijt} = 0$ samples. While there is no hard-and-fast rule that may be applied to determine how much overlap is "reasonable", the condition they include with their `ppml` command is to check whether the mean value of each w^k_{ijt} over $X_{ijt} > 0$ lies between the maximum and minimum values it takes over $X_{ijt} = 0$. Setting aside the more general (and comparably benign) issue of collinearity over *all* X_{ijt}, the only situation where any of our fixed effects would fail this condition would be if a country did not engage in exporting or importing in a given year or if a pair of countries never trade during the sample.[21] Thus, `ppml_panel_sg` drops all observations for pairs of countries who never trade, exporters who don't export anything in a given year, and importers who don't import anything in the given year.[22]

[21] When multiple fixed effects are collinear over the whole sample (as is always the case in this context), these manifest as redundant FOC's which do not affect the existence or uniqueness of a solution for $\widehat{\mathbf{b}}$. Thus, even though one might construct examples where one or more of the fixed effect dummies do not take on both 0 and 1 over each subsample, these scenarios can always be resolved by accounting for general collinearity.

[22] Ultimately, whether or not these observations are dropped or kept does not affect much. What standard Stata commands will do is try to force the conditional mean for these observations to zero, by (wrongly) estimating large, negative values for their associated fixed effects. Stata users should be reassured that, despite this oddity, other estimates are usually fine so long as the main set of non-fixed effect regressors meets the conditions described above.

Table A1: Comparison of Computation Times

	1948-2013		1985-2013		1995-2013		1948-2005		1985-2005		1995-2005	
	HDFE PPML	Standard PPML	HDFE PPML	Standard PPML	HDFE PPML	Standard PPML	HDFE PPML	Standard PPML	HDFE PPML	Standard PPML	HDFE PPML	Standard PPML
5	0:00:05	0:01:04	0:00:02	0:00:11	0:00:02	0:00:08	0:00:04	0:00:54	0:00:02	0:00:06	0:00:02	0:00:05
10	0:00:08	0:06:47	0:00:08	0:00:55	0:00:06	0:00:35	0:00:09	0:04:26	0:00:02	0:00:33	0:00:03	0:00:14
15	0:00:14	0:25:19	0:00:07	0:03:51	0:00:07	0:01:50	0:00:14	0:18:16	0:00:06	0:02:08	0:00:04	0:00:46
20	0:00:34	1:19:04	0:00:12	0:11:14	0:00:13	0:05:08	0:00:27	0:55:35	0:00:09	0:06:13	0:00:05	0:02:22
25	0:00:49	4:07:49	0:00:20	0:25:19	0:00:25	0:11:23	0:00:46	2:56:44	0:00:15	0:14:24	0:00:22	0:06:02
30	0:01:00	10:02:02	0:00:32	1:06:18	0:00:28	0:28:04	0:01:06	7:12:47	0:00:23	0:33:39	0:00:15	0:13:32
35	0:01:29	20:28:41	0:00:41	2:50:11	0:00:32	0:59:14	0:01:20	14:14:27	0:00:33	1:12:04	0:00:39	0:25:30
40	0:01:44	n.c.	0:00:57	6:03:55	0:00:37	2:46:06	0:01:34	n.c.	0:00:41	3:00:26	0:00:27	0:52:57
45	0:02:08	n.c.	0:01:16	13:00:35	0:00:50	6:05:37	0:02:04	n.c.	0:00:57	6:57:02	0:00:34	2:24:56
50	0:02:08	n.c.	0:01:28	21:26:31	0:01:03	10:24:28	0:01:52	n.c.	0:01:01	11:36:54	0:00:38	4:32:44
55	0:01:47	n.c.	0:01:45	n.c.	0:01:38	17:23:56	0:02:05	n.c.	0:01:16	19:37:42	0:00:51	8:01:50
60	0:01:28	n.c.	0:01:39	n.c.	0:02:05	n.c.	0:01:16	n.c.	0:01:25	n.c.	0:00:54	13:00:53
65	0:01:03	n.c.	0:01:45	n.c.	0:01:15	n.c.	0:01:04	n.c.	0:01:34	n.c.	0:00:52	20:52:58
70	0:01:13	n.c.	0:01:29	n.c.	0:01:21	n.c.	0:01:01	n.c.	0:01:31	n.c.	0:01:05	n.c.
75	0:00:58	n.c.	0:01:19	n.c.	0:01:24	n.c.	0:01:09	n.c.	0:01:38	n.c.	0:01:17	n.c.

Notes: This table reports computation times for different sample sizes (both in terms of countries and years considered) for the `ppml`-command of Santos Silva and Tenreyro (2011), in columns labeled 'Standard PPML' and the HDFE `ppml_panel_sg`-command of Zylkin (2017), in columns labeled 'HDFE PPML'. The first column of the table lists the number of countries included in each specification in increasing order. Computation times are given in hh:mm:ss. "n.c." refers to situations where we did not achieve convergence after 24 hours. All estimations performed on a cluster with 2 cores à 3.06MHz and allocated 15GB RAM each. Note that the Stata's maximum number of variables of 32,767 precludes estimations with PPML for example for the full data set form 1948-2013 at the latest for more than 127 countries as one needs to generate $N \times (N-1) + (N \times T \times 2)$ dummies.

Table A2: PPML with Missings as Zero Trade Flows

	All CUs (1)	Disagg. EMU (2)	Disagg. CUs (3)
All CUs	0.153		
	(0.010)***		
	[0.043]***		
	{0.083}*		
EMU		0.0521	0.0489
		(0.010)***	(0.010)***
		[0.043]	[0.042]
		{0.095}	{0.095}
All Non-EMU CUs		0.728	
		(0.026)***	
		[0.110]***	
		{0.180}***	
CFA Franc Zone			-0.126
			(0.100)
			[0.337]
			{0.354}
East Caribbean CU			-0.877
			(0.083)***
			[0.296]***
			{0.296}***
Aussie $			0.384
			(0.119)***
			[0.243]
			{0.226}*
British £			1.060
			(0.035)***
			[0.145]***
			{0.239}***
French Franc			2.096
			(0.063)***
			[0.229]***
			{0.308}***
Indian Rupee			0.170
			(0.147)
			[0.371]
			{0.304}
US $			0.0183
			(0.022)
			[0.066]
			{0.051}
Other CUs			0.766
			(0.053)***
			[0.185]***
			{0.250}***
RTAs	0.159	0.160	0.159
	(0.009)***	(0.009)***	(0.009)***
	[0.040]***	[0.040]***	[0.040]***
	{0.077}**	{0.077}**	{0.076}**
CurCol	0.827	0.630	0.387
	(0.064)***	(0.055)***	(0.047)***
	[0.299]***	[0.250]**	[0.195]**
	{0.291}***	{0.257}**	{0.156}**

Notes: This table reproduces the results from Table 1 of the main text after treating all missing observations in the sample as zeroes. 1,610,165 observations for more than 200 countries for the years 1948 to 2013. Robust standard errors in parentheses. Standard errors clustered by country pairs in brackets. Standard errors clustered by exporter, importer, and year in curly brackets. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

Table A3: OLS Estimation of Different Subsamples

	1948-2013	1985-2013	1995-2013	1948-2005	1985-2005	1995-2005
			All countries			
EMU	0.429	0.444	0.476	0.172	0.176	0.177
	(0.021)***	(0.022)***	(0.028)***	(0.032)***	(0.030)***	(0.037)***
	[0.056]***	[0.049]***	[0.048]***	[0.070]**	[0.059]***	[0.060]***
	{0.149}***	{0.135}***	{0.121}***	{0.158}	{0.140}	{0.120}
All Non-EMU CUs	0.298	0.235	0.301	0.290	0.076	0.167
	(0.025)***	(0.088)***	(0.132)**	(0.026)***	(0.107)	(0.167)
	[0.069]***	[0.171]	[0.217]	[0.066]***	[0.167]	[0.214]
	{0.097}***	{0.183}	{0.224}	{0.091}***	{0.170}	{0.209}
			Industrial countries plus present/future EU			
EMU	-0.010	-0.052	0.043	-0.088	-0.158	-0.074
	(0.021)	(0.022)**	(0.025)*	(0.032)***	(0.031)***	(0.036)**
	[0.055]	[0.044]	[0.041]	[0.069]	[0.059]***	[0.054]
	{0.098}	{0.074}	{0.042}	{0.107}	{0.095}	{0.068}
All Non-EMU CUs	0.537	-0.151	-0.444	0.532	0.300	0.059
	(0.049)***	(0.250)	(0.329)	(0.049)***	(0.275)	(0.302)
	[0.173]***	[0.704]	[0.518]	[0.165]***	[0.643]	[0.483]
	{0.196}***	{0.732}	{0.460}	{0.181}***	{0.644}	{0.440}
			Upper income (GDP p/c $\geq$ \$ 12,736)			
EMU	0.107	0.138	0.163	-0.017	-0.007	-0.085
	(0.026)***	(0.027)***	(0.033)***	(0.037)	(0.035)	(0.041)**
	[0.059]*	[0.054]**	[0.054]***	[0.080]	[0.070]	[0.065]
	{0.103}	{0.094}	{0.099}	{0.123}	{0.104}	{0.108}
All Non-EMU CUs	0.456			0.378		
	(0.138)***			(0.123)***		
	[0.248]*			[0.211]*		
	{0.350}			{0.277}		
			Rich Big (GDP$\geq$\$ 10bn, GDP p/c$\geq$\$ 10k)			
EMU	0.109	0.098	0.094	0.051	0.016	-0.066
	(0.023)***	(0.024)***	(0.029)***	(0.032)	(0.030)	(0.032)**
	[0.055]**	[0.048]**	[0.047]**	[0.071]	[0.058]	[0.049]
	{0.093}	{0.078}	{0.081}	{0.117}	{0.088}	{0.090}
All Non-EMU CUs	1.041			0.990		
	(0.100)***			(0.093)***		
	[0.338]***			[0.300]***		
	{0.263}***			{0.239}***		
			OECD			
EMU	-0.026	-0.034	-0.031	-0.059	-0.077	-0.081
	(0.019)	(0.017)*	(0.021)	(0.027)**	(0.021)***	(0.021)***
	[0.065]	[0.049]	[0.042]	[0.066]	[0.049]	[0.039]**
	{0.084}	{0.053}	{0.033}	{0.073}	{0.047}	{0.041}*
All Non-EMU CUs	1.052			0.998		
	(0.129)***			(0.122)***		
	[0.630]*			[0.586]*		
	{0.668}			{0.621}		
			Present/future EU			
EMU	-0.267	-0.217	-0.037	-0.312	-0.289	-0.099
	(0.024)***	(0.023)***	(0.024)	(0.036)***	(0.032)***	(0.029)***
	[0.069]***	[0.051]***	[0.043]	[0.086]***	[0.070]***	[0.049]**
	{0.112}**	{0.096}**	{0.046}	{0.123}**	{0.125}**	{0.078}
All Non-EMU CUs	0.814			0.736		
	(0.065)***			(0.065)***		
	[0.277]***			[0.272]***		
	{0.417}*			{0.407}*		

Notes: This table reports estimates obtained from linear specifications that correspond to the PPML estimates from Table 2 of the main text. Robust standard errors in parentheses. Standard errors clustered by country pairs in brackets. Standard errors clustered by exporter, importer, and year in curly brackets. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.