2019

Currency Unions and Trade: A PPML Re-Assessment with High-Dimensional Fixed Effects

Mario Larch

Joschka Wanner

Yoto V. Yotov

Thomas Zylkin
University of Richmond, tzylkin@richmond.edu

Recommended Citation

Larch, Mario, Joschka Wanner, Yoto V. Yotov and Thomas Zylkin. "Currency Unions Trade: A PPML Re-Assessment with High-Dimensional Fixed Effects." *Oxford Bulletin of Economics and Statistics* 81, no. 3 (June 2019): 487–510. https://doi.org/10.1111/obes.12283.

Currency Unions and Trade:
A PPML Re-assessment with High-Dimensional Fixed Effects*

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

October 23, 2018

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 associated with the three-way (exporter-time, importer-time, and country-pair) fixed effects currently recommended in the gravity literature have heretofore limited the choice of estimator, leaving an important methodological gap. To address this gap, we introduce an iterative Poisson Pseudo-Maximum Likelihood (PPML) estimation procedure that facilitates the inclusion of these fixed effects for large data sets and also allows for 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. Most importantly, our estimates for both the overall CU effect and the Euro effect specifically are economically small and statistically insignificant. We also document that linear and PPML estimates of the Euro effect increasingly diverge as the sample size grows.

JEL Classification Codes: C13; C23; C55; F14; F15; F33
Key Words: Gravity; EMU effect; Panel data; Poisson; Large data sets

* *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. *Note*: The procedure presented in this paper is implemented in Stata and freely available via ssc (to install, type "`ssc install ppml_panel_sg, replace`") or at https://econpapers.repec.org/software/bocbocode/S458249.htm.

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

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

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

¶ *Affiliation:* University of Richmond. *Address:* Robins School of Business, 1 Gateway Road, Richmond, VA, 23217.

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 analogues. (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), having garnered over 3,100 citations since its original publication in google scholar and 356 citations in Web of Science Core Collection. 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 the "multilateral resistance" constraints implied by theory (Anderson and van Wincoop, 2003; Feenstra, 2004; Baldwin and Taglioni, 2007), and time-invariant pair fixed effects to absorb unobservable barriers to trade (such as bilateral history) and to address the endogeneity of trade policy variables due to time-invariant unobserved bi-

[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), de Sousa (2012), and Campbell (2013) are examples of reactions to Rose's initial finding. Finally, Micco, Stein, and Ordonez (2003), Baldwin and Taglioni (2007), Bun and Klaassen (2007), Berger and Nitsch (2008), Santos Silva and Tenreyro (2010a), Eicher and Henn (2011), Olivero and Yotov (2012), Herwartz and Weber (2013), and Mika and Zymek (2018) specifically investigate the effect of the EMU. Santos Silva and Tenreyro (2010a) and Rose (2017) survey each of these literatures.

lateral heterogeneity (Baier and Bergstrand, 2007).[2] 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 advances that 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 instalment, 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.[3]

Doing their due diligence, Glick and Rose (2016) also experiment with PPML estimation with two-way (exporter-time and importer-time) fixed effects.[4] 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., with pair fixed effects also added to the exporter-time and importer-time fixed effects from the two-way model).[5]

In this paper, we pick up where Glick and Rose (2016) leave off. The main technical

[2] Of course, endogeneity of common currencies may also arise from time-varying bilateral effects. Our investigation does not tackle these sources of selection into currency unions.

[3] Glick (2017) demonstrates that these results are robust to controlling for EU membership and further shows that there is heterogeneity in the trade effects between new and old EMU members.

[4] 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).

[5] Even before Glick and Rose (2016), computational challenges with PPML have been quietly simmering for some time. For example, Bratti, De Benedictis, and Santoni (2014) study the impact of immigrants on trade and note that "[t]he use of [...] the Pseudo Poisson Maximum Likelihood (PPML) estimator [...] clashes with the use of a large set of fixed effects that hamper 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.

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. In order to clearly demonstrate the importance of our methods, we employ the same dataset as Glick and Rose (2016) and we rely on the same theory-consistent gravity model with added pair fixed effects, reflecting the latest developments in the gravity literature noted above. Thus, the differences in our results are driven exclusively by the following two innovations. First (and most importantly), we present an iterative PPML algorithm that specifically addresses the computational burden of the three different types of high-dimensional fixed effects ("HDFEs") that need to be computed to obtain consistent point estimates of Glick and Rose (2016)'s preferred specification.[6] Second, with these consistent point estimates in hand, we take advice from Cameron, Gelbach, and Miller (2011) and Egger and Tarlea (2015) and base our inferences on standard errors that 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.[7]

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, which disappears with the PPML estimator. However, for all CUs *other than the EMU,* we find (as much of

[6]The algorithm we present draws on an earlier method devised by Guimarães and Portugal (2010) for PPML with two-way HDFEs. It was originally programmed by Zylkin (2017) and is available in Stata via ssc (to install, type "`ssc install ppml_panel_sg, replace`") or at https://econpapers.repec.org/software/bocbocode/S458249.htm.

[7]While the current focus is on currency unions and trade, the methods we describe are equally well-suited to a wide variety of other applications that call for the estimation of a gravity model.

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%.

We are not the first to document either a small EMU effect (c.f., Micco, Stein, and Ordonez, 2003; Baldwin and Taglioni, 2007), or, indeed, an insignificant 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.[8] As Glick and Rose (2016) and Rose (2017) rightly point out, using a sample with many countries and years is in principle always the most sensible approach. However, in practice, it is also this preference that contributes to the large difference in estimates. As Santos Silva and Tenreyro (2006) highlight, OLS estimation of the log-linearised gravity model will in general be inconsistent in the presence of heteroscedasticity. We investigate the degree of heteroscedasticity in Glick and Rose (2016)'s data by plotting estimation residuals against expected trade values for different benchmark subsamples split by development status, regions, and size. This analysis reveals that trade flows involving the many smaller, poorer countries needed for a comprehensive sample are noticeably more heteroscedastic than trade flows involving other countries. The inclusion of these countries in Glick and Rose (2016)'s data (and in other similarly large data sets) should therefore be expected to exacerbate the difference between PPML and OLS estimates, a pattern we can confirm by comparing coefficient estimates for different subsamples. We also use the example of the EMU effect to demonstrate that the addition of very small countries that contribute only a tiny portion of world trade can have a noticeable impact on OLS estimates of a currency union even if they are not part of the currency union, whereas PPML estimates will tend to discount the addition of such countries.[9]

[8] A notable exception is Mika and Zymek (2018), 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.

[9] Indeed, Glick and Rose (2016)'s own sensitivity analysis—replicated in Table A2 of our Online Appendix—makes it plain that linear estimates of the EMU effect do depend non-trivially on which non-EMU reference countries are included in the sample. Another recent follow-up study by Campbell and Chentsov

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

Following the latest developments in the gravity literature, we now describe and implement a PPML estimation procedure that can be used to obtain estimates for a large number of exporter-time, importer-time, and exporter-importer ("pair") fixed effects. We also discuss how to resolve the subsequent technical challenge of how to obtain multi-way clustered PPML standard errors in the presence of these fixed effects.

2.1 Estimation Procedure

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. This specification is in line with the best practices for panel gravity estimation recommended by Yotov, Piermartini, Monteiro, and Larch (2016) and has appeared in a number of recent empirical studies on the effects of trade agreements, albeit only with smaller samples.[10] Our goal is to obtain PPML estimates for coefficient vector $\mathbf{b}$ for large samples in the presence of these three high-dimensional fixed effects. To fix ideas, we first write an expression for the corresponding estimate of $\mathbf{b}$, denoted

(2017) adds further emphasis on this point.

[10] See, for example, Dai, Yotov, and Zylkin (2014), Bergstrand, Larch, and Yotov (2015), Anderson, Vesselovsky, and Yotov (2016), Anderson and Yotov (2016), and Heid and Larch (2016) who investigate samples with 13 to 41 regions or countries. Some of these papers facilitate estimation of this specification by further making the simplifying assumption that the pair fixed effect μ_{ij} applies symmetrically in both directions. Thanks to the algorithm introduced in this paper, these compromises are no longer necessary.

by $\widehat{\mathbf{b}}$, in the form of a generalized PPML first-order condition:

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

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

$$\widehat{\lambda}_{it}: \quad Y_{it} - e^{\widehat{\lambda}_{it}} \sum_j \exp\left(\widehat{\psi}_{jt} + \widehat{\mu}_{ij} + \widehat{\mathbf{b}}'\mathbf{w}_{ijt}\right) = 0, \tag{2b}$$

$$\widehat{\psi}_{jt}: \quad X_{jt} - e^{\widehat{\psi}_{jt}} \sum_i \exp\left(\widehat{\lambda}_{it} + \widehat{\mu}_{ij} + \widehat{\mathbf{b}}'\mathbf{w}_{ijt}\right) = 0, \tag{2c}$$

$$\widehat{\mu}_{ij}: \quad \sum_t X_{ijt} - e^{\widehat{\mu}_{ij}} \sum_t \exp\left(\widehat{\lambda}_{it} + \widehat{\psi}_{jt} + \widehat{\mathbf{b}}'\mathbf{w}_{ijt}\right) = 0, \tag{2d}$$

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.[11]

Along with (2a), 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.[12] 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). To do so, first define

$$\Psi_{it} \equiv \frac{Y_{it}/\sqrt{X_{Wt}}}{e^{\widehat{\lambda}_{it}}}, \qquad \Phi_{jt} \equiv \frac{X_{jt}/\sqrt{X_{Wt}}}{e^{\widehat{\psi}_{jt}}}, \qquad D_{ij} \equiv e^{\widehat{\mu}_{ij}}, \tag{3}$$

where $X_{Wt} \equiv \sum_i \sum_j X_{ijt}$ denotes total world trade at time t, to be used as a scaling factor.[13]

[11] 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).

[12] The PPML Hessian is negative definite (Gourieroux, Monfort, and Trognon, 1984). Thus, so long as a solution for $\widehat{b}$ exists—a finer point we discuss further in the Online Appendix—it is guaranteed to be unique. However, the fixed effects in (2b)-(2d) are only determined up to $2N + T$ normalizations, where N and T respectively denote the numbers of countries and time periods.

[13] The utility of this scaling factor is that, as in the analogous system used in Anderson and van Wincoop

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

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

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

$$\mathbf{0} = \sum_i \sum_j \sum_t \left[X_{ijt} - \left(\frac{Y_{it}X_{jt}}{X_{Wt}}\right)\left(\frac{D_{ij}e^{\widehat{\mathbf{b}}'\mathbf{w}_{ijt}}}{\Psi_{it}\Phi_{jt}}\right)\right]\mathbf{w}_{ijt}, \tag{4a}$$

$$\Psi_{it} = \sum_j \frac{X_{jt}/X_{Wt}}{\Phi_{jt}} D_{ij}e^{\widehat{\mathbf{b}}'\mathbf{w}_{ijt}}, \tag{4b}$$

$$\Phi_{jt} = \sum_i \frac{Y_{it}/X_{Wt}}{\Psi_{it}} D_{ij}e^{\widehat{\mathbf{b}}'\mathbf{w}_{ijt}}, \tag{4c}$$

$$D_{ij} = \frac{\sum_t X_{ijt}}{\sum_t \left(\frac{Y_{it}X_{jt}}{X_{Wt}}\right)\left(\frac{e^{\widehat{\mathbf{b}}'\mathbf{w}_{ijt}}}{\Psi_{it}\Phi_{jt}}\right)}. \tag{4d}$$

In (4b) and (4c), Ψ_{it} and Φ_{jt} are analogues 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 (4d), the "pair" fixed effect recommended by Baier and Bergstrand (2007). As with our other fixed effects, we may obtain this last term by equating sums: in this case, pair-wise sums of actual and fitted trade flows over time, as in (4d).

With this system in place, the steps to follow are exactly as outlined in (4a)-(4d). That is: (i) given initial guesses for $\{D_{ij}, \Psi_{it}, \Phi_{jt}\}$, compute a solution for $\widehat{\mathbf{b}}$ using (4a); (ii)-(iii) Update Ψ_{it} and Φ_{jt} using (4b) and (4c); (iv) update D_{ij} using (4d); and (v) return to step

(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. We thus may use $D_{ij} = \Phi_{jt} = \Psi_{it} = 1$ as natural initial guesses for the fixed effects when first solving for $\widehat{\mathbf{b}}$.

(i) with new values for $\{D_{ij}, \Psi_{it}, \Phi_{jt}\}$, iterating until convergence.[14]

2.2 Standard Errors

Of course, computing the point estimates themselves is just one part of the overall high dimensionality problem we must overcome in obtaining inferences. Estimating standard errors also poses a significant technical challenge in this context. Even though in principle we may readily construct from our iterative procedure the complete Hessian matrix associated with our estimation, the usual method of inverting the Hessian to obtain the estimated Poisson variance matrix is likely to be impractical, because of the number and variety of the included fixed effects. Fortunately, recent advances in the related literature offer better alternatives. For the slightly simpler case of a PPML model with two-way HDFEs, Figueiredo, Guimarães, and Woodward (2015) show how the high dimensionality associated with this latter problem may be efficiently discarded by recognizing the variance-covariance matrix of a Poisson regression is proportional to that of an appropriately weighted linear regression, such that the Frish-Waugh-Lovell theorem may then be applied. This same strategy also extends naturally to the case of three-way HDFEs, as we show in our Online Appendix.

More generally, however, our emphasis on standard errors stems from our desire to incorporate assumptions about error correlation (or "clustering") patterns that are most reasonable for our data and model. Again, we draw on recent innovations. In particular, Egger and Tarlea (2015) convincingly argue that standard errors for a panel-data gravity model should allow for simultaneous correlations across all three main dimensions of the panel—exporter, importer, and time—by implementing the "multi-way" clustering methodology first introduced in Cameron, Gelbach, and Miller (2011). Adopting the logic of Egger and Tarlea

[14] One way to obtain $\widehat{\mathbf{b}}$ in step (i) would be to solve for it directly via a nonlinear solver. However, 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), inspired by Guimarães (2016) and further discussed 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) in the high-dimensional fixed effects context.

(2015), it is reasonable to believe there are auto-correlations across time within countries having to do with inertia in trade, as, e.g., bilateral trade responds sluggishly in the short-run to long-run changes in local prices. A similar logic applies to possible cross-sectional dependence within time periods, as general equilibrium price linkages across countries may not fully reflect an idiosyncratic shock to trade at time t.

For added motivation, we also note that clustering simultaneously on i, j, and t actually allows for correlation in the error term within all six possible cluster dimensions $\{i, j, t, it, jt, ij\}$. It thus explicitly nests the typical practice of assuming errors are solely clustered across time within each country-pair ij. For this reason, we will expect multi-way clustering to lead to more conservative inferences, just as in Egger and Tarlea (2015). The details for implementing multi-way clustering in our setting largely follow Cameron, Gelbach, and Miller (2011), requiring only some slight modification to account for the high dimensionality mentioned above. Again, for brevity, we leave these specifics to our Online Appendix.

In sum, our methods allow us to rapidly obtain estimates and flexibly clustered standard errors for our key parameters of interest, even for data structures that would ordinarily be too large for direct estimation to be feasible. 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.[15] 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

[15] 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). The gains in terms of whether and how fast convergence is achieved will obviously vary with the specific soft- and hardware used to implement the procedure. The main takeaway from Table A1 is that there are substantial speed and feasibility gains using our suggested estimation procedure for high-dimensional fixed effects models compared with previously available methods.

may re-produce Glick and Rose's preferred specification with three-way fixed effects either in its original OLS form,

$$\ln X_{ijt} = \lambda_{it} + \psi_{jt} + \mu_{ij} + \boldsymbol{\beta}'\mathbf{z}_{ijt} + \gamma CU_{ijt} + \epsilon_{ijt}, \tag{5}$$

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

$$X_{ijt} = \exp\left(\lambda_{it} + \psi_{jt} + \mu_{ij} + \boldsymbol{\beta}'\mathbf{z}_{ijt} + \gamma CU_{ijt}\right) + \nu_{ijt}, \tag{6}$$

where, in either case, $\mathbf{z}_{ijt}$ denotes a set of non-CU controls (namely dummies indicating the presence of regional trade agreements and current colonial relationships[16]) 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 $E[X_{ijt} - \widehat{X}_{ijt}|\cdot] = 0$. As a consequence, OLS estimates of γ will only be consistent when the OLS error term ϵ_{ijt} is homoscedastic, whereas PPML is consistent under much more general circumstances. Since trade data are generally taken to be heteroscedastic—a supposition we will later confirm—OLS is likely to be biased and inconsistent, with the bias increasing in the degree of heteroscedasticity.[17] For some further motivation, we also note that, unlike with PPML, OLS first-order conditions for the exporter-time and importer-time fixed effects λ_{it} and ψ_{jt} do not re-produce the adding-up constraints (4b) and (4c) typically implied by theory; instead, they equate sums of log trade flows with sums of fitted log flows.

[16]Note this last variable is mainly identified by former colonies gaining independence during the period. It is debatable whether the trade effect of a country's independence is appropriately captured by this dummy. We therefore re-ran our main specifications after dropping all colonies from the sample (reducing the number of observations by 70%). Our results are both qualitatively identical and quantitatively similar.

[17]Santos Silva and Tenreyro (2006, 2011) provide an extensive discussion of this point as well as a comparison study of PPML versus a range of other nonlinear estimators. While PPML implicitly assumes that the variance of ν_{ijt} is proportional to the conditional mean, this assumption only affects the efficiency of the estimator and PPML turns out to generally perform adequately even when this assumption is not met. Fernández-Val and Weidner (2016) and Jochmans (2017) have documented favourable 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.

3.1 Main Results

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 two 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 curly brackets we report multi-way clustered standard errors clustered by exporter, importer, and year, as advocated by Egger and Tarlea (2015).[18]

For all PPML specifications, as suggested by Santos Silva and Tenreyro (2006) and Manning and Mullahy (2001), we perform a Park (1966)-type test for the hypothesis that the multiplicative gravity model can be consistently estimated in the log-linearised form. We obtain a p-value less than 0.001 in all cases, implying that the adequacy of estimating the constant-elasticity model in log-linear form is strongly rejected. As a goodness of fit measure, we also calculate the squared correlation coefficients between observed and predicted dependent variable values (which coincides with the R^2 in the linear case). The R^2 values we obtain are 0.855 for the linear model and 0.987 for PPML, in line with the typical good fit of gravity models.[19]

Our main observations from the estimates themselves 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 multi-way clustered standard errors are used, it also becomes statistically insignificant.

Second, our PPML estimates for the EMU effect in columns (5) and (6) are even less favourable. The estimated EMU coefficients—0.030 and 0.027, respectively—are an order of magnitude smaller than the corresponding linear model estimates shown in columns (2) and

[18]Note that we drop singleton groups (i.e. fixed effects groups with only a single observation) in order to avoid artificially low standard errors due to an overstated number of clusters (see Correia, 2015). This is achieved with the `dropsingletons`-option in the `ppml_panel_sg`-command.

[19]Note that the higher values for PPML are not only driven by its better fit, but also by considering the correlation of the levels of (fitted and observed) trade flows rather than of their logs as in the OLS case.

(3). Furthermore, when clustered standard errors are used, the EMU loses significance.[20]

Third, 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. The negative net EMU effect we observe could be for several reasons. For example, the EMU countries are mainly developed countries that already had comparably strong and stable individual currencies and ~~which~~ were already well-integrated economically. It may be that the types of transaction costs that currency unions alleviate may be more pronounced for countries that are less integrated with one another and/or have weaker currencies to start with. Or, as de Sousa (2012) has argued, it could also be that the importance of a common currency for trade has generally fallen over time due to increased globalization, with the surprising lack of an EMU effect being part of a broader trend.[21] However, it is worth noting that estimates of the effect of non-EMU currency unions are potentially less reliable, because a substantial part of the identifying variation is due to currency union dissolutions that coincided with political events such as warfare, communist takeovers, or colonial independence (Campbell, 2013).

Finally, other individual CU estimates, shown in column (6), are also affected, to varying degrees. In particular, we see the large PPML estimate for non-EMU CUs is driven by the British £, the French Franc, and "other CUs".[22] The finding of very large heterogeneity in the trade effects across different currency unions is in line with previous findings by Eicher and Henn (2011) and Glick and Rose (2016). However, an additional note of caution is

[20] 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.

[21] We investigate how the non-EMU effect changes over time in Section 3.3 and find further evidence that the currency union effect has generally fallen over time.

[22] Note we treat missing observations in the Glick and Rose (2016) data set 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, presented as robustness check later on.

in order (aside from the potential confounding with geopolitical events) for the estimated effects of individual non-EMU currency unions: they tend to be identified based on very little variation in the data. For example, the effects of the French Franc, the East Caribbean Dollar, and the Australian Dollar are estimated based on the variation in bilateral trade flows between only six, four, and three countries, respectively.

In sum, our PPML estimates of the trade-promoting effects of currency unions in Table 1 (and of the EMU and other individual currency unions in particular) are very different than their OLS counterparts. Given the large magnitude of these differences, it is only natural to wonder: Why do the two estimators give us such strikingly different results here? And is there anything about this particular setting—with an unusually large sample of countries—that would lead these estimates to offer such diverging conclusions? We address these questions next.

3.2 Comparing OLS and PPML Estimates for Different Samples

Because the data set from Glick and Rose (2016) we work with is notably very large, there is likely but one main reason behind the difference in the OLS and PPML estimates we obtain. As our Park test results have confirmed, OLS is an inconsistent estimator in this context because it suffers from a heteroscedasticity-induced bias that does not disappear in large samples, whereas PPML can be shown to be consistent.[23]

Thus, to investigate why the difference in estimates is as large as it is, we need to be able to say something about the pattern of heteroscedestacity and why it might induce an especially large bias for this particular sample. Recall that OLS estimates are consistent

[23] Another reason sometimes cited in the literature for why PPML and OLS estimates differ is that the implied moment conditions of OLS estimation make the regressors orthogonal to the difference between the observed and fitted logged trade flows (i.e., $\ln X_{ijt} - \ln \widehat{X}_{ijt}$), whereas the moment conditions used by PPML establish orthogonality to the deviations in levels (i.e., $X_{ijt} - \widehat{X}_{ijt}$). For this reason, Eaton, Kortum, and Sotelo (2013) and Head and Mayer (2013, 2014) conclude that PPML will assign more importance to larger trade flows relative to OLS. However, these implied weighting differences should affect only the efficiency of each estimator in small samples; for large samples, the consistency properties of the two estimators should explain most of the difference in an otherwise correctly specified model.

in the special case where the OLS error term (ϵ_{ijt}) is homoscedastic.[24] A useful way of visualizing how far off the data is from satisfying this assumption is to use what Tukey (1977) calls a "wandering schematic" diagram. To create this diagram—demonstrated in Figure 1—we first group all residuals from the estimation into 20 equal-sized bins, with each bin collecting observations with similar predicted trade values. We then sort these bins from smallest to largest predicted value and construct modified box plots summarizing the distribution of the error term within each bin. As the top left panel of Figure 1 shows—in a visual confirmation of our earlier Park test results—the residuals from our main OLS specification (i.e., from column 2 in Table 1) are clearly not homoscedastic. In particular, both the boxes for each bin (reflecting the first and third quartiles of the distribution) and their associated whiskers (reflecting the adjacent values) grow steadily smaller from left to right as we consider observations with a higher expected trade value, implying that the variance of these residuals is inversely related to the conditional mean across the entire sample.[25] Note that this does not imply that the differences between observed and fitted trade flows get smaller for larger trade flows. It rather implies a decrease in the percentage difference.

To say something about the "degree" of heteroscedasticity using this type of analysis, it is first necessary to offer some concrete benchmarks for comparison. Glick and Rose (2016)'s robustness analysis offers us some standard ways of restricting the sample that are convenient for this purpose. Drawing on Glick and Rose (2016)'s Table 8, the alternative country subsamples we use are: "industrialized countries plus present/future EU" (countries with an IFS code below 200 plus all current EU countries); "upper income" (countries whose GDP per capita exceeds the World Bank "upper income" threshold of $12,736); "rich and

[24]This special case of a homoscedastic error term from the log-transformed model corresponds to the situation described in Santos Silva and Tenreyro (2006) where the conditional variance from the multiplicative model is proportional to the square of the conditional mean. Both estimators should be consistent when this assumption is satisfied; thus, we would expect them to give similar results in this context.

[25]The adjacent values of a distribution, which occur at a distance 1.5 times the inter-quartile range past the nearest edge of the "box", are a standard concept used in data analysis to determine where the tails of the distribution lie.

big" (countries with a GDP per capita of at least $10,000 and/or a GDP exceeding $10 billion), and one sample each for OECD members and for current and future EU members. Figure 2 presents comparisons of OLS and PPML estimates for these various subsamples, along with 90% and 95% multi-way clustered confidence bounds. From these comparisons, it is easy to see that OLS estimates of the EMU effect are only positive and significant for the full sample; otherwise, the OLS estimate is generally much closer to the PPML estimate and is near-zero and statistically insignificant for all subsamples except the "present/future EU" subsample, where it is negative and significant. For non-EMU CUs, we generally observe the PPML estimate is somewhat larger than the corresponding OLS estimate across each of these subsamples, consistent with our results for the full sample. It is also apparent that OLS estimates of both CU variables are generally more sensitive to varying which countries are used, whereas PPML estimates are relatively more stable across different subsamples.

The remaining panels of Figure 1 then show how the heteroscedasticity in the data changes when we restrict the sample to any of these benchmark country groupings. As we observed earlier for the full sample, all the different subsamples feature observations with smaller predicted values exhibiting higher variance than other observations. However, unlike with the full sample, the heteroscedasticity in these subsamples is mainly limited to the observations with smaller expected values versus the rest of the sample. With the exception of the full sample, the righthand-sides of each of these panels are at least close to homoscedastic. As Figure 3 suggests, this is likely because the full sample includes disproportionately more observations with smaller expected trade values, which generally seem to exhibit more variance and more heteroscedasticity than other observations. We can therefore plausibly conclude that moving to a comprehensive sample from any of these benchmark subsamples fundamentally alters the pattern of heteroscedasticity and amplifies the bias affecting the OLS estimates.

Continuing further with our analysis of these subsamples, another result from Figure 2 that draws our curiosity is the close correspondence between the OLS and PPML estimates of the EMU effect across all the samples we consider except for the full sample. The similarity between estimates for the sample of current/future EU members is particularly interesting to us because this sample already includes all the EMU members; the only difference with the larger sample is that the larger sample also adds many non-EU countries to the "reference group" of trade partners against which within-EMU trade is compared in order to identify the EMU effect. Or, more precisely, the addition of more reference group countries mainly affects the estimation via the exporter-time and importer-time fixed effects λ_{it} and ψ_{jt}. Since trade with a smaller trade partner contributes more to the sum of a country's total log trade flows (the key moment used to identify these fixed effects in OLS) than to its total trade in levels (the key moment used in PPML), it's conceivable that at least some of the divergence in estimates reflects the relatively higher importance that OLS places on the many smaller non-EMU countries present in the full sample.

To explore in more detail how differences between linear and PPML estimates evolve with the composition of the reference group, Figure 4 plots estimates from both the linear model and PPML starting with the EU as a whole and then adding one country at a time ranked by 2013 GDP. These estimates reveal that adding more and more (smaller) countries leads to a continually rising OLS estimate—even for the addition of the world's tiniest economies—while the PPML estimates stabilize after the inclusion of around 40 additional countries. Based on the preceding discussion, this divergence is by no means surprising: as we add smaller and smaller countries to the sample, we are tending to make the data more heteroscedastic, thereby amplifying the bias in the OLS estimate. However, we also note that the PPML and OLS estimates shown in Figure 4 are usually influenced in the same direction whenever the next-largest country is added to the sample. As such, the PPML and OLS estimates shown in the figure both seem to agree that trade has fallen between EMU members relative to their trade with the rest of the EU as well as with the six largest non-EU economies (the US, China, Japan, Brazil, and India), which together constitute more than two-thirds of world non-EMU GDP. Intra-EMU trade appears to have risen, however, relative to trade with smaller partners, starting with the seventh largest non-EMU economy (Canada). To interpret our earlier results in light of these patterns, the positive and significant overall OLS estimate of the EMU effect we observe appears to be heavily influenced by the apparent decline in trade between the EMU and the many smaller non-EMU countries in the sample relative to intra-EMU trade. Since PPML naturally discounts the addition of smaller reference group countries, and since this pattern is nowhere to be found for the EMU's most important outside partners, this feature of the data could help explain some of the difference between the PPML and OLS estimates of the EMU effect.[26]

3.3 Other Robustness

To add some final experiments, we consider here the possible role played by zero trade flows (which cannot be included in log-linear models) as well as some possible omitted temporal factors such as lags, trends and anticipation effects. We also examine how the effect of currency unions has changed over time.

Missing and Zero Trade Flows. As noted earlier, in order to obtain the main estimation results, we treated missing observations in the Glick and Rose (2016) data set as missing for both our linear and PPML specifications. Omitting zero trade flows could potentially lead to a sample selection problem biasing our results. If a currency union induces country-pairs to start trading (versus not at all), estimates based on positive trade flows only may lead to a downward bias of the estimated effect of currency unions on trade. While OLS cannot handle zero trade flows without further adjustments, such as adding a small, arbitrary number (see e.g. Linnemann, 1966) or applying the inverse hyperbolic sine transformation (see e.g. Kristjánsdóttir, 2012), they can be directly included into the PPML estimation. In Table 2, we therefore show the results of re-estimating specifications (4)-(6), now treating

[26] To investigate this intuition, we ran our OLS main specification (column (2) of Table 1) with the product of GDPs as weights. Indeed, the results from the weighted OLS regression are more similar to the PPML estimates than the unweighted ones. Most importantly, the estimated EMU coefficient is 0.335 (compared to 0.429 and 0.030 for unweighted OLS and PPML, respectively) and the estimated currency union effect for all other CUs is 0.450 (compared to 0.298 and 0.700 for unweighted OLS and PPML, respectively).

all missing observations as zero trade flows. The results indicate that including zero trade flows hardly affects the estimates. Most importantly, estimating the gravity equation in its multiplicative form still erases the EMU effect. The one small change from our earlier results is that the general CU effect now remains marginally significant even when clustering at the exporter, importer, and year dimension.

Time Periods. Our main specifications rest on the strong assumption that the influence of the currency unions on trade has not changed over the last seventy years. But some recent evidence provided by de Sousa (2012) suggests this may not be a good assumption. In addition, since the EMU does not begin until relatively late in the sample, it is worth investigating whether our estimates change if we use a more recent time series that is more centred on the EMU specifically.

We thus use Table 3 to investigate how the effects of currency unions change over time, both for our full country sample as well as for the benchmark subsamples considered in Glick and Rose (2016)'s Table 8 and in our own Figure 2. The estimates using all years from 1948-2013 for all subsamples are given in the first column; estimates presented in subsequent columns then experiment with how CU effects vary when the sample period begins in either 1985 or 1995 and/or ends in 2005.[27] Consistent with our earlier Figure 2, we find that no single subsample leads to a positive significant effect for the Euro and that the large and positive non-EMU CU effect is robust for all samples with a large time span. For samples beginning in 1985 or 1995, however, the non-EMU effect is always statistically insignificant when multi-way clustered standard errors are used. This latter set of findings lends support to de Sousa (2012)'s earlier observation that the trade-promoting effect of currency unions seems to have weakened significantly over the course of the 20th century. As discussed in de Sousa (2012), one plausible reason for the decreasing effect of currency unions may be increased international economic integration, both in terms of trade and

[27] The corresponding OLS results, which largely replicate Glick and Rose (2016)'s Table 8, are provided in Table A2 in the Online Appendix.

financial globalization. However, we also note that 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.

Quadrennial Data. Because trade flows may require some time to adjust to changes in trade costs, Cheng and Wall (2005) have suggested using intervals of several years rather than yearly data. Following this advice has become general practice in the regional trade agreements literature (see for example Baier and Bergstrand, 2007; Bergstrand, Larch, and Yotov, 2015) and a similar argument conceivably applies to currency unions as well. We therefore re-estimate our main specification based on 4-year intervals instead of using consecutive years. As can be seen from column (1) of Table 4, this hardly affects our point estimates and standard errors, even though we only use about a quarter of the data.[28]

Time Trends. The inclusion of bilateral fixed effects only captures bilateral time-invariant heterogeneity. One step towards capturing bilateral unobservables in a more flexible way is to interact the pair fixed effects with the linear variable year in order to account for pair-specific trends, as suggested in the EMU context by Bun and Klaassen (2007) and for the estimation of regional trade agreement effects by Bergstrand, Larch, and Yotov (2015). The estimates from column (2) of Table 4 demonstrate that the EMU effect continues to be insignificant when these bilateral linear time trends are added, while the effect of all other CUs is about halved.[29] A possible explanation is that when not controlling for bilateral linear time trends, part of the CU effect captures common changes in bilateral unobserved heterogeneity among CU members. The addition of this time trend also requires a further extension to our PPML estimation procedure, which we provide in the Online Appendix.

Lags. As discussed above, our failure to find a significant EMU effect could plausibly

[28]The OLS results for all specifications of Table 4 are presented in Table A3 in the Online Appendix.

[29]These results and all subsequent results in Table 4 continue to use every four years (as in column 1). Our results are similar if we use every year.

be because trade adjusts slowly to the introduction of a common currency rather than all at once. On top of using 4-year intervals, an additional way to explicitly capture sluggish adjustments and phasing-in effects of trade policies is to follow Baier and Bergstrand (2007)'s suggestion of adding lagged explanatory variables. After adding lags of all variables, the contemporaneous EMU effect is still insignificant (see column (3) of Table 4.) However, the lagged value is now small, positive, and statistically significant. The combined EMU (defined as the sum of the contemporaneous effect and the lagged effect) is 0.027, which is again not statistically different from zero ($std.err. = 0.093$). For all other CUs, both the contemporaneous and lagged value are positive, and statistically significant. Their joint size of 0.668 (with $std.err. = 0.188$) is not statistically significantly different from the value from column (1).

Leads. Lastly, in column (4) of Table 4, we estimate specifications including the leads of all variables. A possible interpretation of this experiment is as a placebo test, as we should not see any effect of CUs that are not already in place. Indeed, for the EMU, the lead effect is close to zero and statistically insignificant. For the other CUs the lead is marginally significant, but substantially smaller than the contemporaneous coefficient. The significance of the lead variable could hint at an endogeneity problem or capture anticipation effects or the impact of any other unobserved drivers of trade. Hence, while our analysis demonstrates the effectiveness and empirical relevance of our methods, we again note that estimates of the effects of the non-EMU CUs—both here and in the literature more generally—should be interpreted with at least some caution.

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 with regards to whether the Euro has had a statistically significant effect on trade. Third, we identify a cautionary example where OLS and PPML gravity estimates differ to an especially dramatic degree. We relate this difference to the underlying heteroscedasticity, which renders OLS inconsistent and which increases in the number of small countries included in our sample. Notably, the increasing divergence between estimates for larger samples with more small countries indicates that the computational issues we resolve in this paper would otherwise limit a researcher's choice of estimator precisely when this choice seems to matter most.

References

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

Anderson, J. E., M. Vesselovsky and Y. V. Yotov (2016): "Gravity with Scale Effects," *Journal of International Economics*, 100, 174–193.

Anderson, J. E., and Y. Yotov (2016): "Terms of Trade and Global Efficiency Effects of Free Trade Agreements," *Journal of International Economics*, 99, 279–298.

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.

Bergstrand, J.H., M. Larch, and Y. V. Yotov (2015): "Economic Integration Agreements, Border Effects, and Distance Elasticities in the Gravity Equation," *European Economic Review*, 78, 307–327.

Bratti, M., L. De Benedictis, and G. Santoni (2014): "On the Pro-Trade Effects of Immigrants," *Review of World Economics*, 150(3), 557–594.

Bun, M. J. G., and F. J. G. M. Klaassen (2007): "The Euro Effect on Trade is not as Large as Commonly Thought," *Oxford Bulletin of Economics and Statistics*, 69(4), 473–496.

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.

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," MPRA Working Paper No. 79973.

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.

Cheng, I.-h., and H. J. Wall (2005): "Controlling for Heterogeneity in Gravity Models of Trade and Integration," *Federal Reserve Bank of St. Louis Review*, 87(1), 49–64.

Correia, S. (2015): “Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix,” Unpublished Working Paper.

Correia, S. (2016): “Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator,” Unpublished Working Paper.

Dai, M., Y. V. Yotov, and T. Zylkin (2014): “On the Trade-Diversion Effects of Free Trade Agreements,” *Economics Letters*, 122(2), 321–325.

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.

Eaton, J., S. Kortum and S. Sotelo (2013): “International Trade: Linking Micro and Macro,” in *Advances in Economics and Econometrics: Tenth World Congress*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, pp. 329–370.

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. (2017): “Currency Unions and Regional Trade Agreements: EMU and EU Effects on Trade,” *Comparative Economic Studies*, 59(2), 194–209.

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,” NBER Working Paper No. 21535.

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.

Head, K., and T. Mayer (2013): "What Separates us? Sources of Resistance to Globalization," *Canadian Journal of Economics*, 46(4), 1196–1231.

Head, K., and T. Mayer (2014): "Gravity Equations: Workhorse, Toolkit, and Cookbook," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, vol. 4, chap. 3, pp. 131–195. North Holland, 4 edn.

Heid, B., M. Larch (2016): "Gravity with Unemployment," *Journal of International Economics*, 101, 70–85.

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

Herwartz, H., and H. Weber (2013): "The Role of Cross-Sectional Heterogeneity for Magnitude and Timing of the Euro's Trade Effect," *Journal of International Money and Finance*, 37, 48–74.

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. (2017): "Two-Way Models for Gravity," *The Review of Economics and Statistics*, 99(3), 478–485.

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," 140th Seminar, December 13-15, 2013, Perugia, Italy 163341, European Association of Agricultural Economists.

Kristjánsdóttir, H. (2012): "Exports from a Remote Developed Region: Analyzed by an Inverse Hyperbolic Sine Transformation of the Gravity Model," *The World Economy*, 35(7), 953–966.

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

Linnemann, H. (1966): *An Econometric Study of International Trade Flows*. Holland Publishing, Amsterdam.

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.

Manning, W. G., and J. Mullahy (2001): "Estimating Log Models: To Transform or not to Transform?," *Journal of Health Economics*, 20(4), 461–494.

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 (2018): “Friends Without Benefits? New EMU Members and the ‘Euro Effect’ on Trade,” *Journal of International Money and Finance*, 83, 75–92.

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.

Park, R. E. (1966): “Estimation with Heteroscedastic Error Terms,” *Econometrica*, 34(4), 888.

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.

Tukey, J.W. (1977): *Exploratory Data Analysis*. Reading, MA.

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.

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)***		
	{0.080}***			{0.081}		
EMU		0.429	0.432		0.030	0.027
		(0.021)***	(0.021)***		(0.010)***	(0.010)***
		{0.149}***	{0.149}***		{0.092}	{0.091}
All Non-EMU CUs		0.298			0.700	
		(0.025)***			(0.025)***	
		{0.097}***			{0.172}***	
CFA Franc Zone			0.583			0.137
			(0.100)***			(0.108)
			{0.186}***			{0.307}
East Caribbean CU			-1.637			-1.014
			(0.106)***			(0.081)***
			{0.334}***			{0.319}***
Aussie $			0.389			0.168
			(0.196)**			(0.121)
			{0.248}			{0.282}
British £			0.554			1.004
			(0.034)***			(0.034)***
			{0.101}***			{0.234}***
French Franc			0.874			2.096
			(0.083)***			(0.062)***
			{0.269}***			{0.302}***
Indian Rupee			0.522			0.082
			(0.115)***			(0.149)
			{0.110}***			{0.308}
US $			-0.051			0.014
			(0.063)			(0.022)
			{0.229}			{0.066}
Other CUs			-0.104			0.788
			(0.058)*			(0.052)***
			{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)***
	{0.062}***	{0.061}***	{0.061}***	{0.076}**	{0.075}**	{0.076}**
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)***
	{0.155}*	{0.159}*	{0.170}	{0.288}**	{0.251}**	{0.150}**
N	877,736	877,736	877,736	877,736	877,736	877,736
# of clusters						
exporters	212	212	212	212	212	212
importers	212	212	212	212	212	212
years	66	66	66	66	66	66
(Pseudo-)R^2	0.855	0.855	0.855	0.987	0.987	0.987
Park-Test (p-value)	-	-	-	<0.001	<0.001	<0.001

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. 877,736 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 exporter, importer, and year in curly brackets. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

Table 2: PPML with Missings as Zero Trade Flows

	All CUs (1)	Disagg. EMU (2)	Disagg. CUs (3)
All CUs	0.153		
	(0.010)***		
	{0.083}*		
EMU		0.0521	0.0489
		(0.010)***	(0.010)***
		{0.095}	{0.095}
All Non-EMU CUs		0.728	
		(0.026)***	
		{0.180}***	
CFA Franc Zone			-0.126
			(0.100)
			{0.354}
East Caribbean CU			-0.877
			(0.083)***
			{0.296}***
Aussie $			0.384
			(0.119)***
			{0.226}*
British £			1.060
			(0.035)***
			{0.239}***
French Franc			2.096
			(0.063)***
			{0.308}***
Indian Rupee			0.170
			(0.147)
			{0.304}
US $			0.0183
			(0.022)
			{0.051}
Other CUs			0.766
			(0.053)***
			{0.250}***
RTAs	0.159	0.160	0.159
	(0.009)***	(0.009)***	(0.009)***
	{0.077}**	{0.077}**	{0.076}**
CurCol	0.827	0.630	0.387
	(0.064)***	(0.055)***	(0.047)***
	{0.291}***	{0.257}**	{0.156}**
N	1,610,165	1,610,165	1,610,165
# of clusters			
exporters	213	213	213
importers	213	213	213
years	66	66	66
Pseudo-R^2	0.986	0.987	0.987
Park-Test (p-value)	<0.001	<0.001	<0.001

Notes: This table reproduces the results from Table 1 after treating all missing observations in the sample as zeroes. Robust standard errors in parentheses. Standard errors clustered by exporter, importer, and year in curly brackets. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

Table 3: 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)***
	{0.092}	{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)
	{0.172}***	{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)***
	{0.081}*	{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)
	{0.270}***	{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)***
	{0.073}	{0.052}	{0.037}	{0.066}**	{0.043}**	{0.032}**
All Non-EMU CUs	0.762			0.743		
	(0.130)***			(0.107)***		
	{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)***
	{0.077}	{0.053}	{0.038}	{0.073}	{0.043}**	{0.032}**
All Non-EMU CUs	1.312			1.223		
	(0.059)***			(0.055)***		
	{0.321}***			{0.256}***		
				OECD		
EMU	-0.103	-0.047	-0.027	-0.140	-0.092	-0.069
	(0.012)***	(0.012)***	(0.016)*	(0.016)***	(0.013)***	(0.013)***
	{0.071}	{0.054}	{0.040}	{0.067}**	{0.042}**	{0.032}**
All Non-EMU CUs	1.214			1.171		
	(0.062)***			(0.370)***		
	{0.439}***			{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)***
	{0.099}***	{0.055}	{0.041}	{0.124}***	{0.084}**	{0.063}
All Non-EMU CUs	1.157			1.131		
	(0.054)***			(0.052)***		
	{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). RTAs and CurCol are included in the regressions, but their coefficient estimates are not shown for brevity. Robust standard errors in parentheses. Standard errors clustered by exporter, importer, and year in curly brackets. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

Table 4: PPML with Time Trends, Leads, and Lags

	Intervals	Trends	Lags	Leads
	(1)	(2)	(3)	(4)
EMU	0.022	-0.058	-0.056	0.020
	(0.020)	(0.023)**	(0.025)**	(0.028)
	{0.091}	{0.072}	{0.070}	{0.048}
All Non-EMU CUs	0.701	0.387	0.181	0.556
	(0.050)***	(0.050)***	(0.071)**	(0.067)***
	{0.176}***	{0.133}***	{0.080}**	{0.164}***
RTAs	0.178	0.125	0.077	0.216
	(0.018)***	(0.013)***	(0.018)***	(0.020)***
	{0.086}**	{0.052}**	{0.060}	{0.083}***
CurCol	0.619	0.347	-0.026	0.605
	(0.117)***	(0.096)***	(0.108)	(0.148)***
	{0.300}**	{0.285}	{0.134}	{0.279}**
EMU_{t-4}			0.083	
			(0.026)***	
			{0.035}**	
All Non-EMU CUs_{t-4}			0.488	
			(0.062)***	
			{0.140}***	
RTAs_{t-4}			0.146	
			(0.017)***	
			{0.063}**	
CurCol_{t-4}			0.674	
			(0.118)***	
			{0.232}***	
EMU_{t+4}				-0.031
				(0.026)
				{0.074}
All Non-EMU CUs_{t+4}				0.219
				(0.076)***
				{0.126}*
RTAs_{t+4}				0.019
				(0.021)
				{0.060}
CurCol_{t+4}				-0.040
				(0.144)
				{0.173}
N	221,170	221,170	217,462	196,559
# of clusters				
exporters	212	212	212	211
importers	212	212	212	211
years	17	17	16	16
Pseudo-R^2	0.987	0.995	0.987	0.986
Park-Test (p-value)	<0.001	<0.001	<0.001	<0.001

Notes: Column (1) of this table reproduces the results of column (5) of Table 1 but using the data in four year intervals. In addition, we add bilateral linear time trend in column (2) and lags and leads in columns (3) and (4), respectively. Robust standard errors in parentheses. Standard errors clustered by exporter, importer, and year in curly brackets. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

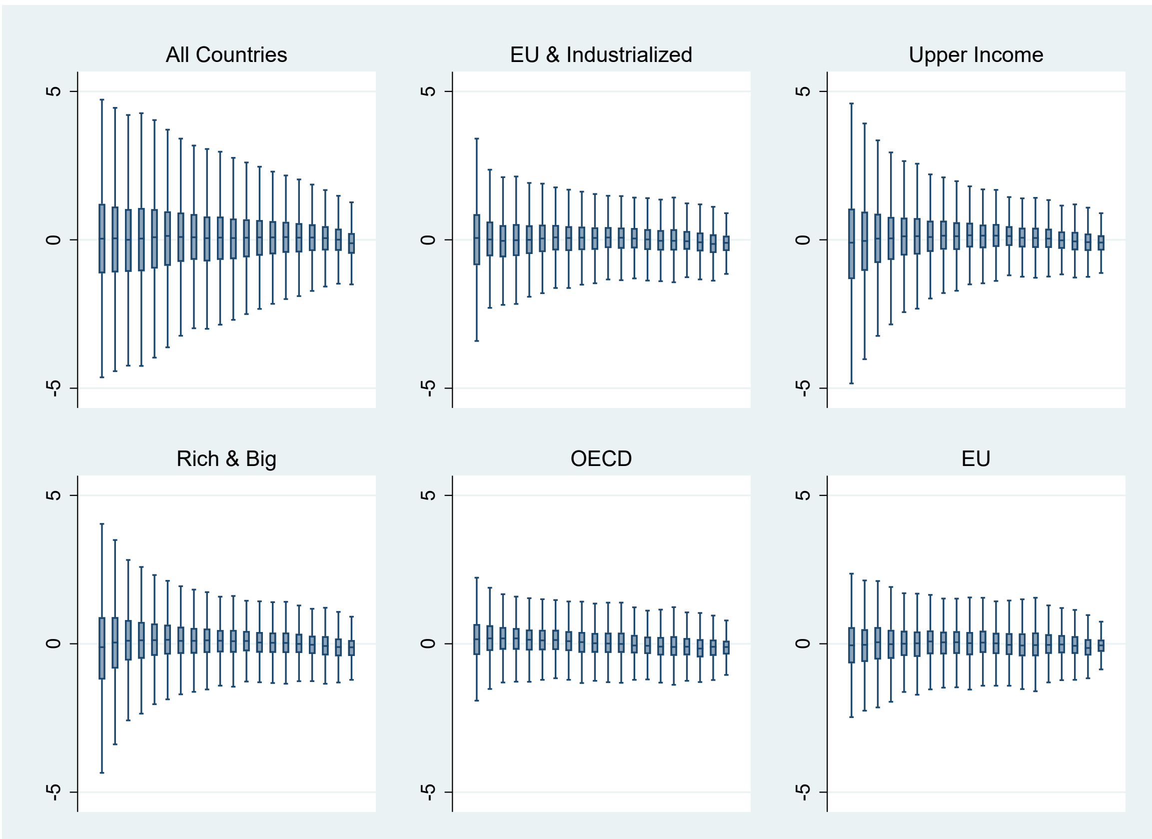

Figure 1: Visualizing Heteroscedesticity in the Data: OLS Residuals vs. Predicted Log Trade Flows, Binned by Size.

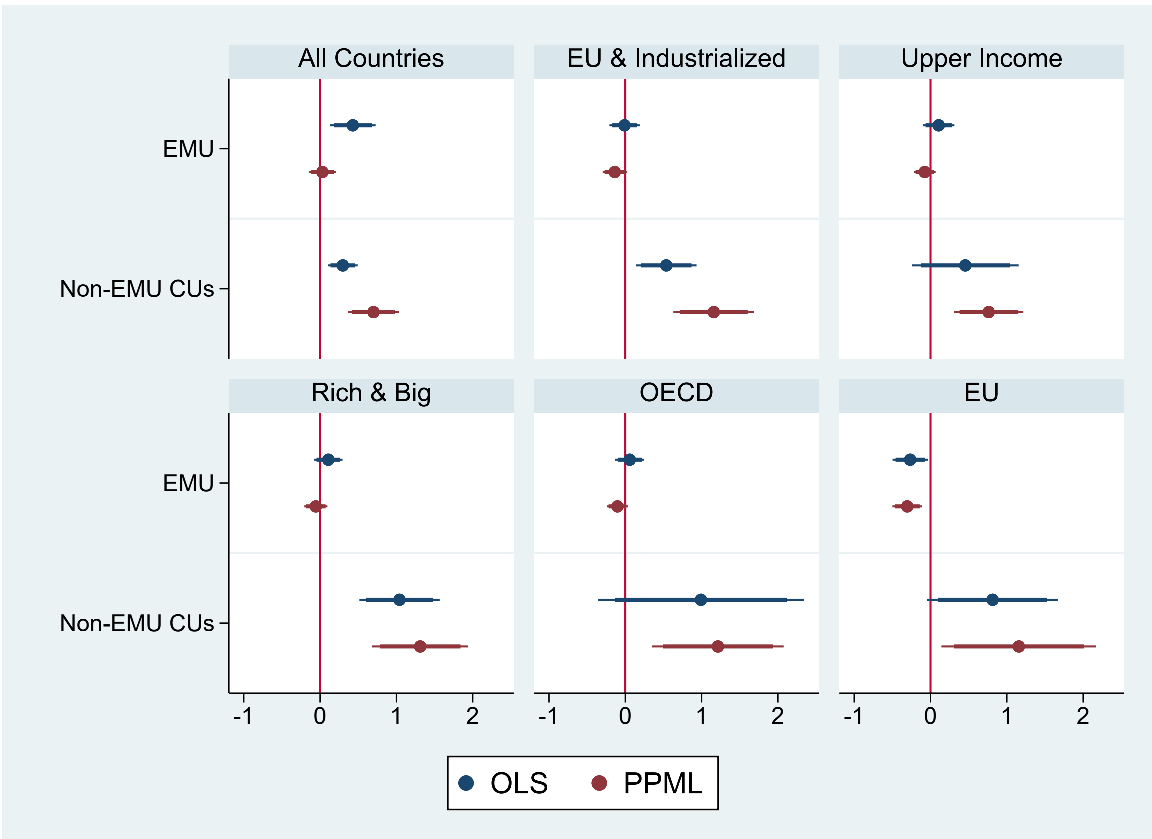

Figure 2: Comparing OLS and PPML Results Across Different Country Samples (with 90% and 95% multi-way clustered confidence bounds).

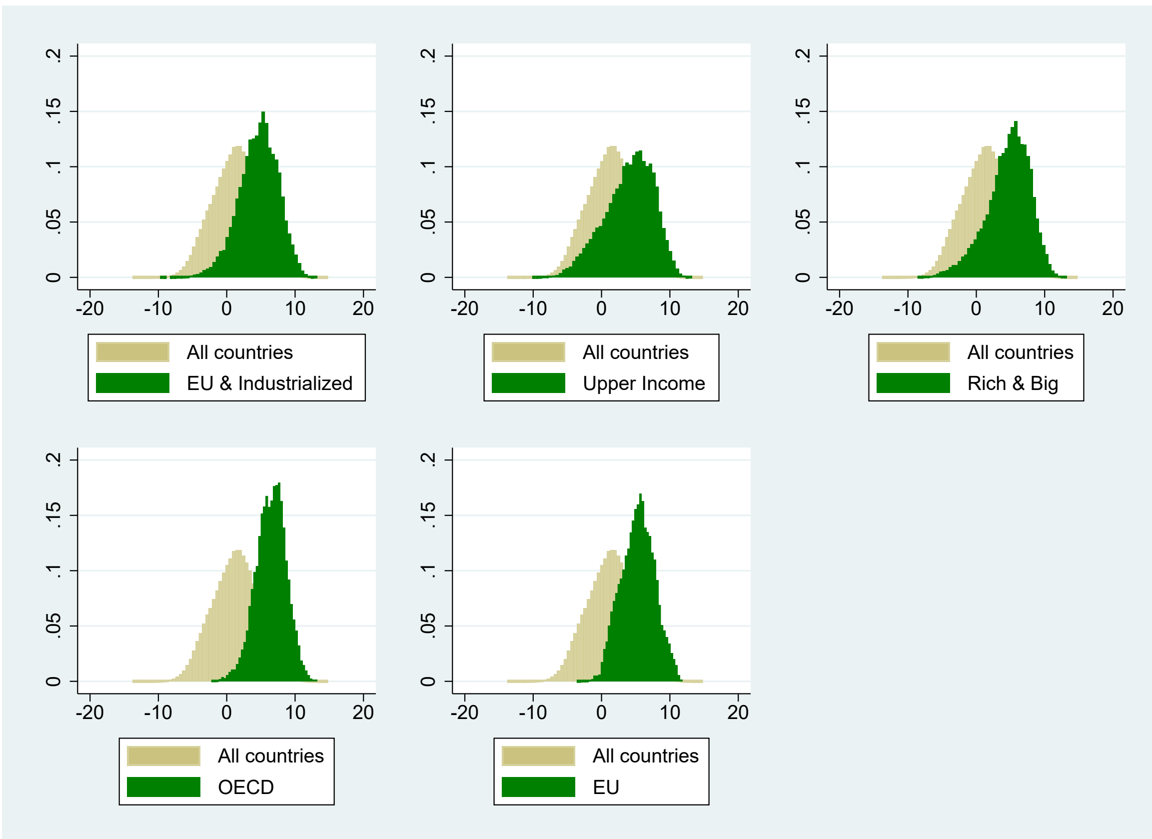

Figure 3: Density Plots of Expected Log Trade Values, by Subsample.

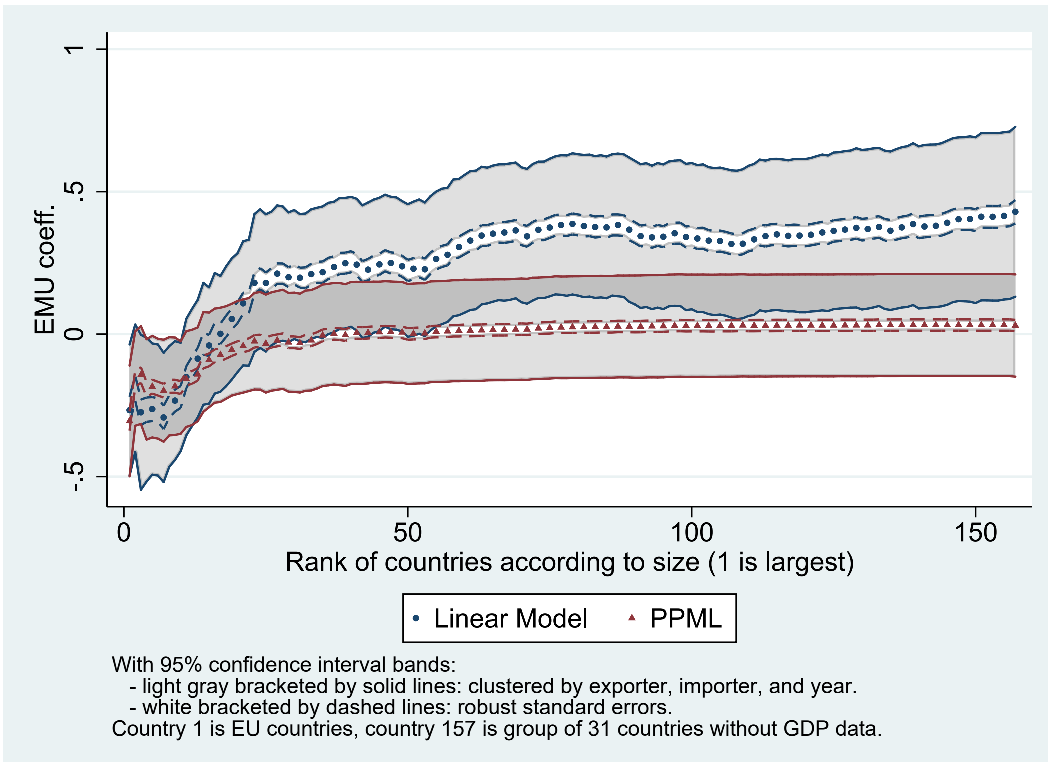

Figure 4: The Effect of Varying the Reference Group of non-EMU countries on EMU Coefficient Estimates

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

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` or implement the proposed procedure in other software packages such as Matlab or R.[30] 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.[31] What must be added here are the intermediate steps needed to compute Ψ, Φ, and D, which follow from (4b)-(4d). 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

[30] We use Stata because it is the most widely used software by trade economists running gravity regressions. However, the procedure described here can be easily implemented in other software packages as well.

[31] 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).

approach versus solving the first-order condition for $\widehat{\mathbf{b}}$ exactly each time through the loop.[32]

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 (2010). 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-

[32]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).

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 Frish-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,

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

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

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

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 (2010), 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.[33]

[33] 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 (2010) for a simple example of a model with non-collinear regressors that does not have a solution.

With multiple high-dimensional fixed effects, implementing the checks favoured by Santos Silva and Tenreyro (2010) 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 $X_{ijt} > 0$.

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 (2010) 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 trades during the sample.[34] Thus, `ppml_panel_sg` drops all observations for pairs of countries who never trade, exporters who do not export anything in a given year, and importers who do not import anything in the given year.[35]

Time trends. For time trends, let α_{ij} be the time trend coefficient and let $t = 0, 1, 2, 3...$

[34]When multiple fixed effects are collinear over the whole sample (as is always the case in this context), these manifest as redundant FOC's that 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.

[35]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.

be the time trend itself. The estimating equation is now given by:

$$X_{ijt} = \exp\left(\lambda_{it} + \psi_{jt} + \mu_{ij} + \alpha_{ij}t + \mathbf{b}'\mathbf{w}_{ijt}\right) + \nu_{ijt}. \tag{A5}$$

The PPML first-order condition for α_{ij} is

$$\sum_t \left(X_{ijt} - \widehat{X}_{ijt}\right) t = 0,$$

which again amounts to a summation of actual and fitted flows, only this time multiplied by the trend at time t (which we are here taking to be one and the same).

Now suppose we have an initial guess value α_{ij}^0 for the time trend and we want to obtain the next value in a converging sequence α_{ij}^1. To obtain α_{ij}^1 we may write:

$$\sum_t \left(X_{ijt} - \widehat{X}_{ijt}^0 e^{d\alpha_{ij}t}\right) t = 0, \tag{A6}$$

where $d\alpha_{ij} = \alpha_{ij}^1 - \alpha_{ij}^0$ is the change in α_{ij} from one iteration to another and $\widehat{X}_{ijt}^0$ are current fitted values. The idea is that when the α_{ij}'s converge, $d\alpha_{ij} = 0$ implies that the first-order condition is satisfied. We want to obtain a new value for $d\alpha_{ij}$ based on (A6) that will allow us to update $\alpha_{ij}^1 = \alpha_{ij}^0 + d\alpha_{ij}$, but unfortunately (A6) is nonlinear in $d\alpha_{ij}$ and cannot be solved analytically. Thus, we instead derive a first-order Taylor Series expansion around $d\alpha_{ij} = 0$:

$$\sum_t \left(X_{ijt} - \widehat{X}_{ijt}^0\right) t - d\alpha_{ij} \sum_t \widehat{X}_{ijt}^0 t^2 = 0, \tag{A7}$$

since $f'(0)$ in this case is $-\sum \widehat{X}_{ijt}^0 t^2$. We then solve (A7) to obtain $d\alpha_{ij}$, update $\alpha_{ij}^1 = \alpha_{ij}^0 + d\alpha_{ij}$, iterate on all other first-order conditions, and repeat until convergence. The system of equations we now need to solve to obtain standard errors is:

$$\sum_j \widehat{X}_{ijt}\left(w_{ijt}^k - \widetilde{\lambda}_{it}^{k*} - \widetilde{\psi}_{jt}^{k*} - \widetilde{\mu}_{ij}^{k*} - \widetilde{\alpha}_{ij}^{k*}t\right) = 0 \quad \forall i, t, \tag{A8a}$$

$$\sum_i \widehat{X}_{ijt}\left(w_{ijt}^k - \widetilde{\lambda}_{it}^{k*} - \widetilde{\psi}_{jt}^{k*} - \widetilde{\mu}_{ij}^{k*} - \widetilde{\alpha}_{ij}^{k*}t\right) = 0 \quad \forall j, t, \tag{A8b}$$

$$\sum_t \widehat{X}_{ijt}\left(w_{ijt}^k - \widetilde{\lambda}_{it}^{k*} - \widetilde{\psi}_{jt}^{k*} - \widetilde{\mu}_{ij}^{k*} - \widetilde{\alpha}_{ij}^{k*}t\right) = 0 \quad \forall i, j, \tag{A8c}$$

$$\sum_t \widehat{X}_{ijt}\left(w_{ijt}^k - \widetilde{\lambda}_{it}^{k*} - \widetilde{\psi}_{jt}^{k*} - \widetilde{\mu}_{ij}^{k*} - \widetilde{\alpha}_{ij}^{k*}t\right) t = 0 \quad \forall i, j, \tag{A8d}$$

which is again our weighted within-transformation exercise from before only with the last set of equations representing the first-order conditions from a linear time trend.

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 labelled 'Standard PPML' and the HDFE `ppml_panel_sg`-command of Zylkin (2017), in columns labelled '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. The exact speed gains as well as the applicable constraints depend on specific soft- and hardware used to implement the procedure. The speed of the `ppml`-command can be improved by using re-scaled trade flows instead of their original values. Nonetheless, the results in this table indicate clear speed and feasibility improvements of the HDFE `ppml_panel_sg`-command.

Table A2: 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)***
	{0.149}***	{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)
	{0.097}***	{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)**
	{0.098}	{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)
	{0.196}***	{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)**
	{0.103}	{0.094}	{0.099}	{0.123}	{0.104}	{0.108}
All Non-EMU CUs	0.456			0.378		
	(0.138)***			(0.123)***		
	{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)**
	{0.093}	{0.078}	{0.081}	{0.117}	{0.088}	{0.090}
All Non-EMU CUs	1.041			0.990		
	(0.100)***			(0.093)***		
	{0.263}***			{0.239}***		
			OECD			
EMU	0.058	-0.001	-0.027	0.035	-0.038	-0.077
	(0.017)***	(0.015)	(0.019)	(0.023)	(0.018)**	(0.018)***
	{0.093}	{0.053}	{0.032}	{0.086}	{0.048}	{0.039}*
All Non-EMU CUs	0.991			0.947		
	(0.129)***			(0.120)***		
	{0.664}			{0.615}		
			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)***
	{0.112}**	{0.096}**	{0.046}	{0.123}**	{0.125}**	{0.078}
All Non-EMU CUs	0.814			0.736		
	(0.065)***			(0.065)***		
	{0.417}*			{0.407}*		

Notes: This table reports estimates obtained from linear specifications that correspond to the PPML estimates from Table 3 of the main text. RTAs and CurCol are included in the regressions, but their coefficient estimates are not shown for brevity. Robust standard errors in parentheses. 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 with Time Trends, Leads, and Lags

	Intervals (1)	Trends (2)	Lags (3)	Leads (4)
EMU	0.431	0.361	0.225	0.055
	(0.042)***	(0.047)***	(0.057)***	(0.073)
	{0.169}**	{0.128}**	{0.148}	{0.141}
All Non-EMU CUs	0.348	0.076	0.238	0.275
	(0.051)***	(0.065)	(0.081)***	(0.073)***
	{0.106}***	{0.116}	{0.087}**	{0.098}**
RTAs	0.414	0.053	0.207	0.325
	(0.019)***	(0.022)**	(0.025)***	(0.028)***
	{0.084}***	{0.067}	{0.109}*	{0.115}**
CurCol	0.321	-0.015	-0.076	0.341
	(0.069)***	(0.074)	(0.101)	(0.103)***
	{0.154}*	{0.135}	{0.132}	{0.111}***
EMU_{t-4}			0.141	
			(0.068)**	
			{0.105}	
All Non-EMU CUs_{t-4}			0.075	
			(0.073)	
			{0.124}	
RTAs_{t-4}			0.358	
			(0.027)***	
			{0.103}***	
CurCol_{t-4}			0.358	
			(0.089)***	
			{0.119}***	
EMU_{t+4}				0.280
				(0.063)***
				{0.137}*
All Non-EMU CUs_{t+4}				0.127
				(0.082)
				{0.114}
RTAs_{t+4}				0.183
				(0.026)***
				{0.084}**
CurCol_{t+4}				-0.136
				(0.117)
				{0.080}
N	221,170	221,170	217,462	196,559
# of clusters				
exporters	212	212	212	211
importers	212	212	212	211
years	17	17	16	16
R^2	0.864	0.914	0.865	0.866

Notes: Column (1) of this table reproduces the results of column (2) of Table 1 but using the data in four year intervals. In addition, we add bilateral linear time trend in column (2) and lags and leads in columns (3) and (4), respectively. Robust standard errors in parentheses. Standard errors clustered by exporter, importer, and year in curly brackets. * $p < 0.10$, ** $p < .05$, *** $p < .01$. See text for further details.

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.

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.

CORREIA, S. (2016): "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator," Unpublished Working Paper.

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.

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.

NELDER, J. A., AND R. W. M. WEDDERBURN (1972): "Generalized Linear Models," *Journal of the Royal Statistical Society. Series A (General)*, 135(3), 370–384.

SANTOS SILVA, J. M. C., AND S. TENREYRO (2010): "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.

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.