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,

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

must be adjusted to take into account the fact that each $\tilde{w}^{k*}_{ijt}$ 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

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

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 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]

with non-collinear regressors that does not have a solution.

[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.

Time trends. For time trends, let α_{ij} be the time trend coefficient and let $t = 0, 1, 2, 3...$ 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^0_{ij} for the time trend and we want to obtain the next value in a converging sequence α^1_{ij}. To obtain α^1_{ij} we may write:

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

where $d\alpha_{ij} = \alpha^1_{ij} - \alpha^0_{ij}$ is the change in α_{ij} from one iteration to another and $\widehat{X}^0_{ijt}$ 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 $\alpha^1_{ij} = \alpha^0_{ij} + d\alpha_{ij}$, 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 $d\alpha_{ij} = 0$:

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

since $f'(0)$ in this case is $-\sum \widehat{X}^0_{ijt} t^2$. We then solve (A7) to obtain $d\alpha_{ij}$, update $\alpha^1_{ij} = \alpha^0_{ij} + d\alpha_{ij}$, 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:

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

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

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

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

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.