Grav$_{it}$y*

Rodolphe Desbordes[a] Markus Eberhardt[b]

[a] *SKEMA Business School – La Défense Campus, Pôle Universitaire Léonard de Vinci, Esplanade Mona Lisa, Courbevoie, 92916 Paris La Défense Cedex, France*

[b] *School of Economics, University of Nottingham, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, England*

January 23, 2019

Abstract: A large empirical literature has estimated the effects of international agreements on trade flows, but these estimates may not be very informative for individual policymakers as they are averages of potentially very different impacts across country-pairs. We investigate this issue using quarterly data from 20 advanced economies over the past 55 years. Our empirical implementation adopts a Poisson pseudo-maximum likelihood estimation technique and estimates a reduced-form gravity relationship at the pair-level using new insights from the common factor model literature to capture multilateral resistance and globalisation effects. We find an average effect of entering regional trade agreements of just over 20%. However, we demonstrate that this average hides substantial heterogeneity across countries, with direct consequence for heterogeneity across time periods. This heterogeneity is also prevalent in the common currency effect. Beyond these findings, our empirical approach offers solutions to issues currently discussed in frontier gravity model research.

Keywords: trade gravity model, panel data, heterogeneity, multilateral resistance, regional trade agreement, heterogeneity

JEL classification: F13, F14, C23

*We are grateful to Alberto Behar, Simone Bertoli, Eric Bond, Thierry Mayer, Marcelo Olarreaga, João Santos Silva, Halis Murat Yildiz and seminar/session participants at the SIRE Workshop on 'Theory and Estimation of Gravity Equations' at Strathclyde University, CERDI in Clermont-Ferrand, the International Association for Applied Econometrics conference at Queen Mary University of London, the European Trade Study Group conference in Munich, the Nottingham GEP workshop, the IMF Research Department and the 3rd Panel Data Workshop at the University of Amsterdam for useful comments and suggestions on earlier versions of the paper. The usual disclaimers apply. Financial support from SIRE and the UK ESRC (Economic and Social Research Council) [Eberhardt: Grant number ES/K008919/1] is gratefully acknowledged. Correspondence: rodolphe.desbordes@skema.edu

1 Introduction

A tremendous number of papers have estimated the effects of international agreements on trade. The meta-analysis of Head and Mayer (2014) suggests that regional trade agreements (RTAs) and currency unions (CUs) have moderate positive impacts. But how should we interpret the effects derived from pooled estimations of typical trade gravity models? Is the estimated effect *common* to all countries in the sample, or does it represent an *average*? These questions have obvious policy relevance, in that an average effect is arguably at best meaningless and at worst seriously misleading for policy decisions *in an individual country* if this average hides substantial differences in the impact *across countries*. Furthermore, if there is cross-country heterogeneity of effects, does it even make sense to compare the estimates of studies diverging in their time window of observation? The extensive literature on the effects of RTAs and CUs on international trade is largely silent on these issues.[1] In this paper, we proceed in two steps to illustrate why the standard gravity model of empirical trade should be amended to arrive at 'Grav$_{it}$y' where the implications of variations in i and t are seriously considered.[2]

In a first step we relax the assumption that the gravity relationship is *common across countries*. We introduce a simple empirical implementation for 'heterogeneous gravity' which builds on the now-standard PPML estimator and illustrate its results compared with 'pooled gravity' implementations for a quarterly panel of 20 OECD countries from 1960Q1 to 2014Q4. We follow in the footsteps of the recent panel time series literature and we adopt a common factor structure to model multilateral resistance, globalisation effects, and other forms of spatial dependence.[3] Our implementation thus adopts the PPML estimator for analysis at the country pair-level, proxying multilateral resistance/unobserved common factors using cross-section averages of the observed trade determinants — this approach adopts a Boneva and Linton (2017) extension of the seminal Pesaran (2006) linear common correlated effects (CCE) estimator to the generalised linear model. We compare the average results from this heterogeneous CCE-PPML with those from a conventional pooled PPML with pairwise fixed

[1]For recent reviews see Baldwin, et al. (2008), Santos Silva and Tenreyro (2010), Head and Mayer (2014), and Piermartini and Yotov (2016).

[2]Admittedly, the former variation in our own notation should be across ni, not i, though surely only a pedant would insist on consistent notation in this case.

[3]We are not the first to recognise the suitability of a multi-factor error structure to model multilateral resistance (Serlenga and Shin, 2007; Camarero, Gomez, and Tamarit, 2013; Mastromarco, Serlenga, and Shin, 2016), though ours is the first approach to adopt a PPML estimator and to allow for heterogeneity in the trade policy coefficients across country pairs.

effects and with pairwise and directional-time fixed effects. Our findings show that the RTA effect is somewhat more moderate in magnitude (around 20%) in our heterogeneous than in these pooled implementations (around 55% and 34%, respectively). Studying the country-specific results we show that while some countries appeared to benefit significantly (especially the UK and Ireland), others had to resign themselves to more modest or even negative RTA effects (Greece, and especially Mexico). The CU effect is insignificant and mildly negative for this set of countries but this average hides vast differences across member countries, with Finnish and Irish exports boosted significantly whereas German and Dutch exports suffered.

In a second step we investigate how country-pair heterogeneity translates into the variability in estimated trade policy effects *over time* by taking advantage of the time series observations of our data and adopting a rolling window of estimation of length 20 years (80 quarters). Since countries in our sample (with a single exception) do not leave RTAs or CUs, this empirical setup directs the attention away from some 'average trade policy effect' over time and across countries toward the effect of *specific* economies entering into an RTA or CU over the 20 years following entry. We find that the estimated effects can substantially vary across windows, depending on which country-pairs contribute to the identification of RTA and CU effects. In other words, the time window of observation is not an innocuous choice.

The existing literature closest to this work are the papers by Novy (2013), Baier, Bergstrand, and Clance (2018; BBC), Chen and Novy (2018; CN) and Baier, Yotov, and Zylkin (2019; BYZ), which like us worry about the heterogeneity of trade policy effects across countries.[4] Novy (2013) develops a micro-founded gravity equation which allows for heterogeneous trade costs across country pairs and estimates 'translog gravity' using OLS in a cross-section of OECD country pairs. CN build on this translog gravity model and study the currency union effect in a vast panel dataset covering virtually all goods trade in the post-WWII period. Their fixed effects models study the theoretically-derived link between import shares and the magnitude of the CU effect, which they show to be inversely related. BBC separately consider the intensive and extensive margins of trade and introduce trade policy heterogeneity via interaction terms with standard time-invariant gravity variables (distance, language, colony, etc.) in a fixed effects OLS model. BYZ analyse manufacturing trade, allow for the

[4]Additional work adopting a spatial econometric framework (Kelejian, Tavlas and Petroulas, 2011; Behrens, Ertur and Koch, 2012; Koch and LeSage, 2015) is limited to the cross-section.

presence of internal trade and globalisation effects, use PPML, and estimate country-pair specific RTA effects.[5]

Our empirical approach complements and extends the contributions of the aforementioned papers in various ways. First, it is accepted but not yet widely recognised in empirical research, that in the presence of heterogeneous coefficients, the fixed effects estimates are uninformative weighted averages of country-specific slopes (Juhl and Lugovskyy, 2014) and the fixed effects estimator is frequently not a consistent estimator of the average treatment effect (Chernozhukov, Fernández-Val, Hahn and Newey, 2013), or the average individual slope (Campello, Galvao, and, Juhl, 2018). Hence, a natural progression of gravity research ought to be towards relaxing the assumption of common slopes for all regressors. If desired, once the country-pair coefficients have been estimated, common patterns can be investigated and various aggregations can be applied. Second, multilateral resistance is a major source of spatial dependence but it is not the only one. Kapetanios, Mastromarco, Serlenga, and Shin (2017) highlight that including country-time fixed effects does not remove the bias induced by heterogenous global factors. There is hence a need to capture globalisation effects and other forms of spatial dependence in a more flexible way. Third, in the current empirical setup adopted by researchers, the impact of country-specific factors on international trade cannot be estimated. While some researchers have recently advocated identification through the addition of internal trade flows (Beverelli, Larch, Keck and Yotov, 2018), such a strategy requires access to these illusive data.[6] An alternative strategy which does not rely on internal trade flows but still accounts for multilateral resistance terms is needed. Fourth, heterogeneity over time has traditionally not been considered, casting doubts on the external validity of some findings in the literature. The approach presented in this paper can address all these issues and can therefore be seen at pushing forward the frontier of gravity research.

The remainder of this paper is structured as follows: Section 2 discusses our dataset and sources, in Section 3 we provide the empirical model and introduce our empirical implementation. Results are presented in Section 4 before some concluding remarks.

[5] Our paper also speaks to recent work by Bas, Mayer, and Thoenig (2017) who employ firm-level data from France and China to estimate bilateral-specific aggregate trade elasticities.

[6] Such a requirement is even more challenging if one desires to investigate the determinants of other flows, such as financial flows.

2 Data

We want to relax the assumption of common trade policy effects across countries. We also want to study the consequences of doing so for the estimation of effects at different time periods. In order for these analyses to be feasible we need a substantial number of time series observations. One option would have been to delve into the rich historical data available (e.g. Fouquin and Hugot, 2016; Barbieri and Keshk, 2017), but a concern over extreme events — the Great War and World War II — as well the narrative split of 19th and 20th century economic history into discrete regimes — the First Era of Globalisation, the Interwar period, and the Second and Third Eras of Globalisation after 1945 — make for a poor fit with our present-day policymaker in mind. Instead, we opted to increase the frequency of the data by using quarterly bilateral trade flows for 20 OECD countries available through the IMF Direction of Trade (DOTS) database.

We further employ matched quarterly GDP data from the OECD and trade policy information from the NSF-Kellogg Institute Data Base on Economic Integration Agreements (EIA) compiled by Scott Baier and Jeffrey Bergstrand. The trade policy data is annual, so that we adopt the convention that policy implementation is in the first quarter of the year — at least with regard to the European economies which dominate our sample this is aligned with agreements typically coming into effect on January 1st. Our final sample comprises 20 OECD countries from 1960Q1 to 2014Q4. Sample makeup and descriptive statistics are provided in an appendix.

Table 1 offers some insights into the evolution of trade policy arrangements in our sample: the 55 sample years divide into discrete policy regimes, of which there are eleven for RTAs (thus ten changes to the number of country pairs with RTAs) and four for common currency arrangements. The first 13 years of the sample saw no change in the network of RTAs. The sharpest increase follows in 1973, when the share of country pairs with RTAs jumps by more than 20%. Regimes thereafter are much shorter in length, and economic integration is more gradual, with the exception of a 10% increase in 1986. Over 55 sample years roughly 14% of country pairs were always in an agreement (and thus do not contribute to the identification of a policy effect), 53% entered and 33% did not enter into an RTA. Our sample is thus well-suited to study the effect of RTAs with substantial variation across countries and time. The same cannot be said for CU effects, where an idiosyncratic episode aside the sample merely captures the (two-phased) introduction of the ECU/Euro.

Figure 2 demonstrates how the different policy regime changes affect our rolling window analysis of RTAs (top panel) and CUs (bottom panel), though our discussion here focuses on the former. The grey shading highlights a number of specific 20-year windows, of which there are a total of 36 (Window #1 ends in 1979, Window # 36 in 2014); the starting points of the solid horizontal lines indicate various regime changes, e.g. in 1973 Ireland (IRL) entered an RTA with other European economies, similarly for Finland (FIN), Greece (GRC) and Spain (ESP) over the next decade; a solid line indicates the period for which this specific increase in the RTA network contributes to the identification of the RTA effect, e.g. Window #13 starting in 1972 will just be able to still identify the effect of the Irish entry since it contains one year (four observations) of pre-entry and 19 years (76 observations) of post-entry data for Ireland. Since the windows shift one year rather than one quarter at a time Window #14 cannot capture the Irish RTA effect.[7] Hence for rolling windows ending after 1992 in this example we use dashes instead of a solid line to signal that they cannot identify the Irish RTA effect. All rolling window results below are presented with reference to the end year of the 20-year sample window.

3 Empirical Strategy

3.1 The Gravity Equation

We assume a gravity relationship in the panel as comprising a subset of more general gravity models in which bilateral trade of exporter i to destination market n at time t is given by[8]

$$X_{nit} = \frac{Y_{it}}{\Omega_{it}} \frac{X_{nt}}{\Phi_{nt}} \phi_{nit} \qquad \text{where} \quad 0 \leq \phi_{nit} \leq 1. \tag{1}$$

X_{nit} is a trade flow from an exporter to a destination market, Y_{it} is the value of production for the exporter and X_{nt} the value of expenditure in the destination market n on all source countries — the latter two are typically proxied by

[7]The choice of annual instead of quarterly shifts helps to smooth out the extreme estimates one would obtain from one and 79 observation(s) in and out of regime, respectively.

[8]This exposition builds on the discussion in Santos Silva and Tenreyro (2006), and Piermartini and Yotov (2016) for the gravity model, and Cameron and Trivedi (1998) for Poisson regression.

GDP in the exporter and destination markets, respectively.[9] ϕ_{nit} captures the 'bilateral accessibility' for destination n and exporter i: this contains trade costs between the two markets and any variable which may affect these, including time-variant and invariant, observed and unobserved factors.

A major development in gravity modelling over the past decade following the seminal contribution by Anderson and Van Wincoop (2003) is the recognition that the conditional trade between destination n and exporter i (the conditions being the 'bilateral accessibility') cannot be viewed in isolation from the set of opportunities open to importer n in sourcing goods from exporters other than i and the relative access exporter i has to destinations other than n.[10] The multilateral resistance variables for each actor in the exchange of goods are defined in terms of the bilateral accessibility-weighted exporter capabilities and importer characteristics respectively: exporter $i = 1, \ldots, N-1$ and for importer $n = 1, \ldots, N-1$, and $i \neq n$ let

$$\Omega_{it} = \sum_{\ell=-i} \frac{\phi_{\ell it} X_{\ell t}}{\Phi_{\ell t}} \qquad \Phi_{nt} = \sum_{\ell=-n} \frac{\phi_{n\ell t} Y_{\ell t}}{\Omega_{\ell t}} \tag{2}$$

where $-n$ and $-i$ signify that these magnitudes are not defined in reflexive terms and thus exclude destination n and exporter i from the respective MLR terms of the trading relationship between these two markets.

For our derivation of the empirical gravity model we assume a stochastic version of equation (1)

$$X_{nit} = \frac{Y_{it}}{\Omega_{it}} \frac{X_{nt}}{\Phi_{nt}} \phi_{nit} \eta_{nit} \tag{3}$$

where η_{nit} is an error factor with $E[\eta_{nit} | Y_{it} \Omega_{it}^{-1} X_{nt} \Phi_{nt}^{-1} \phi_{nit}) = 1$.

A very general empirical equivalent to equation (3) allows for flexible unknown parameters on the observable mass and accessibility variables:

$$\begin{aligned} X_{nit} = \; & exp[\beta_{nit}^{i} \, ln(Y)_{it} + \beta_{nit}^{n} \, ln(X)_{nt} + \gamma_{nit} \, ln(\phi)_{nit} \\ & + ln(\Omega)_{it} + ln(\Phi)_{nt}] \, \eta_{nit}, \end{aligned} \tag{4}$$

[9] If X_{ni} is merchandise trade then theory-consistency dictates Y_i to be gross production of traded goods (not simply value-added/GDP) and X_n the apparent consumption of goods, production plus imports minus exports (Head and Mayer, 2014).

[10] This network of dependencies is formalised by econometricians as the deviation from the assumption of 'cross-section weak dependence' (Andrews, 2005; Chudik, Pesaran, and Tosetti, 2011; Chudik and Pesaran, 2015b).

where superscripts are used to identify the coefficient of exporter versus importer GDP/expenditure. This specification is the most general empirical model possible where all unknown parameters on observable variables (β^i, β^n, γ) vary at the pair-level and over time. We demonstrate how this relates to the models employed in the existing literature by adopting parameter restrictions.

3.2 Model Restrictions and Pooled Estimation

A first restriction, adopted in the vast majority of studies, is to assume the gravity model estimates are fixed *across time*:

$$\begin{aligned} X_{nit} &= exp[\beta^i_{ni}\, ln(Y)_{it} + \beta^n_{ni}\, ln(X)_{nt} + \gamma_{ni}\, ln(\phi)_{nit} \\ &\quad + ln(\Omega)_{it} + ln(\Phi)_{nt}]\, \eta_{nit}, \end{aligned} \tag{5}$$

A notable exception here is the study by Klasing, Milionis, and Zymek (2015) who allow for time-variation in three distinct regimes over their long panel from 1870 to 2005. In our analysis of the post-WWII period we first follow the bulk of the literature and estimate policy effects which are specified as time-invariant; later-on we *partly* relax the assumption of fixed parameters over time by adopting 20-year rolling regression windows.

Conventionally, further restrictions in the panel gravity literature are to assume common parameters on the observable variables ($\beta^i_{ni} = \beta^i$, $\beta^n_{ni} = \beta^n$, $\gamma_{ni} = \gamma$). Pairwise fixed effect are added to capture trade policy endogeneity (Baier and Bergstrand, 2007). Additionally including exporter-time and importer-time fixed effects can capture the MLR terms (Olivera and Yotov, 2012; Fally, 2015; Piermartini and Yotov, 2016) — with implications for β^i and β^n (see below).

$$\begin{aligned} X_{nit} &= exp[\beta^i\, ln(Y)_{it} + \beta^n\, ln(X)_{nt} + \gamma\, ln(\phi)_{nit} + \delta_{ni} \\ &\quad + \omega_{it} + \psi_{nt}]\, \eta_{nit}, \end{aligned} \tag{6}$$

where ψ_{nt} and ω_{it} represent the respective directional-time interactions. Thus δ_{ni}, ϕ_{nt} and ω_{it} are the unknown parameters estimated on the various fixed effects in the reduced-form panel gravity model; in our empirical analysis we present one set of results adopting this setup but ignoring the MLR terms (PPML-1), and another accounting for them via the directional-time fixed effects (PPML-2), implemented by using the now-standard (pooled) PPML estimator: this can address concerns over potential biases induced by heteroskedasticity in trade flows and the presence of zero trade flows (Santos Silva and Tenreyro, 2006), as well as exploit the convenient properties of this estimator to capture MLR in combination with directional fixed effects (Olivera and Yotov, 2012; Fally, 2015). In the implementation of PPML-2 the economic mass terms drop out since they are swept into the directional-time fixed effects.

3.3 Heterogeneous Parameter Estimation

A practical difficulty arises if a more flexible specification for the observable variables such as that laid out in equation (5) on the one hand (β_{ni}^{i}, β_{ni}^{n}, γ_{ni}), and the aforementioned recommended practice to capture MLR terms on the other are to be combined: for pairwise heterogeneity in the economic mass and policy variable parameters it is most convenient to estimate equation (5) *separately* for each country pair, which trivially also addresses policy endogeneity since δ_{ni} is captured by an intercept in this pair-level time series regression. However, we cannot include fixed effects for *all exporters and all importers* in an equation for a single importer-exporter pair, let alone interacted with time dummies. Existing studies in the literature which allow for trade policy heterogeneity employ interaction effects (e.g. Baier, Bergstrand and Clance, 2018) or maintain a sets of fixed effects in a PPML model but use pair-specific trade policy dummies (Baier, Yotov and Zylkin, 2019). An alternative approach is to draw on recent insights from the panel time series literature (e.g. Pesaran, 2006; Bai, 2009) and to employ a multi-factor error structure to capture the unobserved MLR terms.

First, we bring the error factor on the inside of the exponential function so as to capture all unobservables u_{nit}:

$$\begin{aligned} u_{nit} &= \delta_{ni} + \omega_{it} + \psi_{nt} + ln(\eta)_{nit} \\ &\equiv \delta_{ni} + \omega_{it} + \psi_{nt} + \varepsilon_{nit}. \end{aligned} \tag{7}$$

Next, we suggest that the dimensionality problem of dealing with a large number of unknown parameters (ω_{it} and ϕ_{nt} add up to 2×20 countries $\times 220$ time series observations $= 8,800$ directional-time dummies) in a country-pair equation of at most 220 time series observations can be solved by imposing more structure on these unobservables, or rather, by studying the makeup of the structural MLR terms these parameters/fixed effects are meant to capture: in the macro panel econometric literature it is widely acknowledged that large datasets of macro variables can be represented by a small number of common factors with heterogeneous factor loadings. For instance, in the forecasting literature Stock and Watson (2002) have shown that 149 macroeconomic time series can be reduced to two or three principal components. Furthermore, Bai (2009) discusses a number of macro-, microeconometric and finance applications where common factors can be employed to model unobserved time-varying heterogeneity in a tractable way.[11]

In the case at hand we posit that the economic mass and accessibility variables along with the trade flows and MLR terms are all driven by a small number of common factors with heterogeneous factor loadings across country-pairs. Thus, we argue that a small number of unobserved common factors $\boldsymbol{f}_t$, each with country pair-specific factor loadings $\boldsymbol{\varphi}_{ni}$, can account for the evolution of trade flows, GDP, etc. In the notation introduced above:

$$\begin{aligned} u_{nit} &= \delta_{ni} + \omega_{it} + \psi_{nt} + \varepsilon_{nit} && (8) \\ &\approx \delta_{ni} + \boldsymbol{\varphi}'_{ni}\boldsymbol{f}_t + \varepsilon_{nit}, && (9) \end{aligned}$$

where an 'approximate factor structure' is represented by a set of common factors $\boldsymbol{f}$, and the associated factor loadings $\boldsymbol{\varphi}$.

How 'significant' is the presence of unobserved common factors in these data, in particular in the evolution of exports and GDP? Before we turn to the linear common correlated effects approach to tackle a multi-factor error structure and its extension to the generalised linear model in the following paragraphs we briefly provide some evidence for the pervasiveness of common factors in the macro panel data at hand. In Figure 1 we illustrate the evolution of common factors and their relation to the observed data: in panel (a) we plot the first and second principal components of exports and GDP in the destination country extracted from the 380 country pairs in our sample.[12] Together these account for almost 90% and 99% of the variation in the respective observed variables. In panel (b) we present the results from two correlation exercises using 20-year rolling windows (the x-axis indicates the end year of each window): first, we

[11] The multifactor error structure has been applied to capture country-specific time-varying total factor productivity (Eberhardt and Teal, 2013a,b; Eberhardt and Presbitero, 2015) or knowledge spillovers in the analysis of sector-level production functions augmented with sectoral R&D stock (Eberhardt, Helmers, and Strauss, 2013).

[12] A principal components analysis can be understood as a data reduction technique whose purpose is to find normalised linear combinations of a set of variables which retain most of the information provided by these variables. When variables are highly correlated, the first linear combination (component) will capture most of the total variance.

take the first principal component of exports and correlate this with GDP in the destination country; the light-coloured box plots are all between .8 and 1, indicating the common factor extracted from the trade data is a sound predictor for GDP evolution. Second, we take the first principal component of destination GDP and correlate this with observed exports; here the correlations are somewhat more moderate, but with the exception of the final three sample windows the median correlation coefficient is in excess of .8 and the inter-quartile range is above .6. We conclude from this exercise that not only do trade and GDP follow a strong factor structure but it appears that there is substantial overlap between the latent forces driving trade on the one hand and those driving GDP on the other.

The insight gained in the recent panel time series literature from this setup in the *linear* regression case is that the unobserved common factors can be captured by observables, either via principal component analysis (Bai, 2009) or using cross-section averages of the dependent and independent variables (Pesaran, 2006). We briefly develop the latter approach (the 'common correlated effects' or CCE estimator) and provide a mathematical indication of the intuition at play — this is for ease of illustration, since it will only be a small step in terms of implementation from the linear model to a generalised linear one (Boneva and Linton, 2017).

For simplicity we assume a double-index of t for the time series and i for the cross-section — we can think of the latter as a placeholder for the country pair as the unit of analysis like in the gravity model, i.e. $i \equiv ni$. Let

$$
\begin{aligned}
y_{it} &= \alpha_i + \beta_i x_{it} + u_{it} \qquad u_{it} = \lambda_i f_t + \varepsilon_{it} &\quad (10)\\
x_{it} &= \delta_i + \phi_i f_t + \varrho_i g_t + e_{it}, &\quad (11)
\end{aligned}
$$

where ε and e are white noise processes. This setup indicates that the single regressor x is driven by the same common factor f as the dependent variable y, albeit with different parameters.[13] In addition there are some factors $\boldsymbol{g}$ which only drive x but not y. This setup is very standard in the macro panel literature and we refer to the studies in footnote 11 for additional details on factor evolution, parameter distributions, etc. It is clear from equations (10) and (11) that x is endogenous and that failing to account for the presence of the unobserved

[13]In a setup with multiple factors we can instead assume that only a subset of $\boldsymbol{f}$ 'overlaps' between the two equations.

common factors will lead to omitted variable bias.[14]

Pesaran's (2006) approach posits that the unobserved common factor f can be captured by the cross-section averages of y and x provided the cross-section dimension of the panel is not too small.[15] A simple algebraic derivation can provide the intuition for the mechanism at work: take the cross-section average of equation (10) and solve it for the common factor f:

$$\bar{y}_t \;=\; \bar{\alpha} + \bar{\beta}\bar{x}_t + \bar{\lambda} f_t \quad \Leftrightarrow \quad f_t = \bar{\lambda}^{-1}(\bar{y}_t - \bar{\alpha} - \bar{\beta}\bar{x}_t), \tag{14}$$

where the bars indicate cross-section averages and the error term disappears since $\bar{\varepsilon} = 0$ by assumption. Next, plug the expression for f back into the original equation (10)

$$\begin{aligned} y_{it} &= \alpha_i + \beta_i x_{it} + \lambda_i \bar{\lambda}^{-1}(\bar{y}_t - \bar{\alpha} - \bar{\beta}\bar{x}_t) + \varepsilon_{it} && (15)\\ &= \alpha_i - \lambda_i \bar{\lambda}^{-1}\bar{\alpha} + \beta_i x_{it} + \lambda_i \bar{\lambda}^{-1}\bar{y}_t - \lambda_i \bar{\lambda}^{-1}\bar{\beta}\bar{x}_t + \varepsilon_{it} \\ y_{it} &= \varpi_i + \beta_i x_{it} + \zeta_i \bar{y}_t + \vartheta_i \bar{x}_t + \varepsilon_{it}, && (16) \end{aligned}$$

where we reparameterize in the last line. Thus the unobserved common factor f can be captured by the cross-section averages of y and x, while the heterogeneous impact of f across i can be captured by estimating equation (16) separately for each panel member — the principle extends to multiple factors. A Mean Group estimator following Pesaran and Smith (1995) captures the central tendency of the panel and provides a convenient comparison with alternative pooled empirical models:

$$\hat{\beta}^{MG} = N^{-1}\sum_{i=1}^{N}\hat{\beta}_i \tag{17}$$

Inference for the Mean Group estimates is based on a simple nonparametric

[14] Solving the x equation for f and plugging this into the y equation yields

$$\begin{aligned} y_{it} &= \alpha_i + \beta_i x_{it} + \lambda_i \phi_i^{-1}(x_{it} - \delta_i - \psi_i g_t - e_{it}) + \varepsilon_{it} && (12)\\ &= \alpha_i - \lambda_i \phi^{-1}\delta_i + (\beta_i + \lambda_i \phi_i^{-1})\, x_{it} - \lambda_i \phi_i^{-1}\psi_i g_t - \lambda_i \phi_i^{-1} e_{it} + \varepsilon_{it} \\ &= \eta_i + \theta_i x_{it} + \nu_{it}, && (13) \end{aligned}$$

where in the final line we simply reparameterise. Crucially, unless $\lambda_i \phi_i^{-1} = 0$ we can see that β_i is unidentified. The asymptotic bias will be a function of the (relative) 'strength' of the factors in their impact on y and x in panel member i.

[15] Additional robustness of this approach to nonstationary factors, structural breaks, additional spatial dependence, among other aspects, is discussed in Kapetanios, Pesaran, and Yamagata (2011), Chudik, Pesaran and Tosetti (2011), Chudik and Pesaran (2015a) and Westerlund (2018).

variance estimator (Pesaran and Smith, 1995; Pesaran, 2006):[16]

$$\widehat{\sum}^{MG} = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{\beta}_i - \hat{\beta}^{MG})(\hat{\beta}_i - \hat{\beta}^{MG})'. \tag{18}$$

Boneva and Linton (2017) extend the above setup from the linear model to a setting where the outcome variable is discrete. They show that the CCE approach can be applied to a probit model by including only the cross-section averages of the observed regressors: under the assumption that the unobserved factors are contained in the span of the cross-section averages of the regressors they derive asymptotic results for the large T, large N case as well as the consistency and asymptotic normality of the Mean Group estimator of the individual-specific estimates.

The same principle can be applied to a generalised linear model: in our case we begin by assuming the exponential mean function incorporating a multi-factor error structure[17]

$$E[y_{it}|x_{it}, \boldsymbol{f}_t] = \mu_{it} = exp[\alpha_i + \beta_i x_{it} + \boldsymbol{\lambda}_i' \boldsymbol{f}_t]. \tag{19}$$

The Poisson Maximum Likelihood (PML) estimator assumes that the distribution of y given x (and the factors) is Poisson (i.e. a count variable), but it is widely recognised that the data generating process is not required to be Poisson for this estimator to be consistent (Cameron and Trivedi, 1998; Santos Silva and Tenreyro, 2006), in which case it is referred to as a Poisson Pseudo Maximum Likelihood (PPML) estimator. Our implementation uses the estimator in the single time series, namely of the pair-wise gravity equation with exponential mean function, where the common factors are replaced by cross-section averages of the regressors, which we refer to as CCE-PPML:

$$\begin{aligned} \mu_{nit} &= exp[\alpha_{ni} + \beta_{ni}^{i}\, ln(Y)_{it} + \beta_{ni}^{n}\, ln(X)_{nt} + \gamma_{ni}\, ln(\phi)_{nit} \\ &\quad +\delta_{ni}\, \overline{ln(Y)}_t + \kappa_{ni}\, \overline{ln(\phi)}_t] \qquad \forall\, ni,\, n \neq i. \end{aligned} \tag{20}$$

In our case the accessibility term is replaced by dummies for RTAs and CUs.[18]

[16] In practice we follow the standard in the literature and employ robust means (Hamilton, 1992) to reduce the effect of outliers.

[17] We thank Lena Boneva and Oliver Linton for sharing the rough derivations for a CCE-augmentation in this more general case.

[18] We know that the dependent variable, conditional on the regressors, can be heteroskedastic, and serially correlated, and that the PPML estimator maintains its consistency in the presence of

Our empirical estimates below are thus based on cross-section average-augmented time series PPML regressions at the country-pair level which are subsequently averaged following the Mean Group principle; inferential statistics for the CCE-PPML Mean Group estimate are computed using the variance estimator in (18): Boneva and Linton (2017) have shown the simple nonparametric variance estimator still applies in the generalised linear model setup.

Estimation of the CCE-PPML model can be seen as an improvement on current practice for two reasons. First, the presence of common factors, with heterogeneous loadings across country-pairs, offers a flexible way to account not only for multilateral resistance, but also so-called globalisation effects, and other forms of spatial dependence. Second, using common factors as proxies for spatial dependence allows the estimation of country-specific variables. On the other hand, current fixed effects gravity models do not deal with all types of spatial dependence, and do not permit the estimation of country-specific variables in the absence of data on internal trade.

4 Results

4.1 Full time series results

We begin with the results presented in Table 2 where we use three PPML estimators to analyse the full time series data from 1960 to 2014, adopting different data frequencies: quarterly, annual (selecting Q3 observations) and tri-annual observations (selecting Q3 observations every third year), though the latter is not used for the CCE-PPML where the time series dimension matters a lot. The first set of results in columns [1]-[3] are for a Pooled PPML with pairwise fixed effects (henceforth PPML-1), results in [4]-[6] are for a Pooled PPML where in addition to pairwise fixed effects exporter-time and importer-time (directional-time) fixed effects are included (henceforth PPML-2), and results in [7] and [8] are for the Heterogeneous CCE-PPML. It is notable that the data frequency does not affect our empirical results in any substantial way, meaning we can focus on estimates for the quarterly data, since these data will also be used in the rolling window analysis below.

Estimates for PPML-1 suggest large effects for both RTAs and CUs, in excess of

serial correlation if the exponential conditional mean is correctly specified (Cameron and Trivedi, 1998: 226) — the latter is the requirement for any PML or PPML estimator.

50% and close to 30%, respectively, though it should be noted that the standard errors are quite large (especially relative to those in the PPML-2 and PPML-CCE models). We know that excluding directional-time fixed effects from these models is akin to ignoring the forces captured by the MLR terms (Olivera and Yotov, 2012; Piermartini and Yotov, 2016) and these estimates therefore cannot account for the presence of third-market effects. The PPML-2 results indicate the magnitude of this bias, with RTA effects now a more moderate 34%. The CU effect is here estimated to be around -3%, though the difference in statistical significance across the different data frequencies in [4]-[6] suggests we should not take this estimate at face value. The CCE-PPML estimates (robust means across pair-specific estimates) suggest an RTA effect of 23% and also indicate a negative CU effect albeit insignificant in the quarterly data case. Here of course the estimates on the trade policy variables are only available if the country-pair experience any change, i.e. the two countries entered into an agreement during the sample period, which is the case in 202 and 92 pairs for RTAs and CUs, respectively. Estimates for Origin and Destination GDP are qualitatively unchanged if we estimate the average for these 202 or 92 pairs, only.

Table 1 only presents pooled and averaged estimates, but a major advantage of the Heterogeneous CCE-PPML approach is that we can address the question how representative this average is for individual countries in the sample. Figures 3 and 4 provide some answers to this question. The country-specific effect in these figures is split into the effect on exports in Panel (a) and imports in Panel (b).

In the RTA results in Figure 3 we exclude Australia, Canada, Japan, and the United States since their averages would be based on four or fewer observations. All other countries have more RTA partners, between 9 (GBR among others) and 18 (MEX), but given the small samples involved we provide two sets of confidence intervals for 90% (light grey) and 68% (dark grey). Mexico is thus the only non-European country in this exercise. Our estimates for exporters confirm that there are statistically significant differences between some of the country estimates and hence that an average effect (as highlighted with the thick dashed line) is misleading for policymaking in individual countries. All Central and Northern European countries (except the EEA members Sweden and Norway) have positive significant RTA effects, with Ireland (mean estimate 69%) and the Brexit-bound United Kingdom (58%) topping the rankings. Only Greece and Mexico have negative estimates, which in the case of the latter (-41%) are large and statistically significant. On the import side Spain has experienced a vast RTA effect (91%), with Germany (31%), Denmark (40%) and once again Ireland (43%) also benefitting (statistically significant estimates), while Mexican imports suffered (-20%).

The CU effect in Figure 4 is limited to the Euro-area countries in the sample since the Anglo-Irish episode in the 1970s merely provides for 2 estimates. The effects on exports are negative significant in both Germany and the Netherlands, while Finland and Ireland saw large and extremely large effects, respectively. Other countries in the graph had statistically insignificant effects. On the import side no country had statistically significant effects though the large positive estimate for Portugal is borderline significant at the 10% level.

Focusing on the i in Grav$_{it}$y, this exercise has revealed substantial heterogeneity in trade policy effects across countries. We now turn to study how country-pair heterogeneity can lead to time heterogeneity t when the window of observation changes.

4.2 Rolling window results

Country-pair heterogeneity naturally implies that the estimated impacts of RTAs and CUs can greatly differ depending on which country-pairs contribute to identification. We illustrate this by adopting a 20-year rolling window of analysis, meaning that the effect of a specific country-pair entering into an RTA or currency union is only identified for the 19 years following entry. Figure 5 presents the results from PPML regressions using the same three implementations as above: (a) the pooled PPML with pairwise fixed effects (PPML-1), (b) the pooled PPML with pairwise and directional-time fixed effects (PPML-2), and (c) the heterogeneous CCE-PPML. Along the x-axis we indicate the corresponding window *end-year* for each estimate, from 1981 to 2014.[19] The solid line in each plot indicates the estimated RTA effect along with its 95% confidence interval (shaded area), the short-dashed line the CU effect (limited to the experience of the creation of the EMU) with 95% confidence interval (shaded area) — in case of the CCE-PPML these are robust mean estimates across the pairwise estimates for each window (following Pesaran and Smith, 1995; Hamilton, 1992). In all three plots we indicate the full time series result for RTA effects from Table 2 using a horizontal dashed line.

[19] We drop the first two window estimates for both the RTA and CU effects to avoid overemphasizing the very short-run following introduction of RTAs and the Euro/ECU.

A first point to note is that the full sample estimate is a sound representation of an average over time for the PPML-2 and CCE-PPML in panels (b) and (c), but not for the PPML-1 in panel (a), where it instead seems to capture the *maximum* effect of the early 2000s. The 'evolution' of the RTA effect as revealed by the rolling window analysis is also distinct across estimators: for the PPML-1 it rises to almost .5 (equivalent to 65%) by 2004, though interrupted by a slump from 1988 to 1992, and then declines quite sharply in the aftermath of the global financial crisis, ending up at around .1 (11%); the PPML-2 describes a more steady rise to a similar maximum of .5 (65%), much less interrupted by the 1988 slump, but declines earlier from 1999 to 2008, subsequently levelling out at zero; finally, the CCE-PPML describes a sharp slump in 1985, then stays virtually flat until 2004 at around .25 (28%), before declining to zero and below in the remainder of the sample period. The effects described are statistically significantly different from zero except in the more recent period for the PPML-2 and CCE-PPML.

Figure 6 provides a more detailed investigation of the evolution of RTAs in their relation to the CCE-PPML rolling window estimates indicated with the solid or dashed line (estimate transformed into percent): the robust mean estimate after 2004 is no longer statistically significantly different from zero (10% level), which is highlighted by the dashed instead of solid line. All RTAs involving only European/EEC countries are printed in blue, those including other countries are printed in black; note that 'entry' into an agreement here means just that, while 'exit' refers to the point in time when the country or countries mentioned do no longer contribute to the identification of the RTA effect since their 'entry' took place 19 years ago — to reiterate, no country in this sample left an RTA. The figures highlighted in circles along the mean estimates are the number of country-pairs which identify the RTA coefficient in the respective rolling window, for instance the rolling window ending in 1981 had 114 (out of 380) country pairs which *entered* into an RTA in the period 1962-1981.

It can be seen that the characteristic evolution of the mean estimate for RTAs can be matched to changes in the sample: the first slump in the CCE-PPML evolution takes place between 1985 and 1986 as Spain enters the EEC and Finland expands its EEC-agreements to seven additional partners. These new entrants along perhaps with Greece (which entered in 1981) represent an expansion beyond the European core of countries which brought down the average RTA effect to a still-respectable 25% by the late 1980s. Although after 1992 78 bilateral agreements within Europe signed in 1973 no longer help to identify the RTA effect the mean estimate therafter remains relatively stable at around 25%. Similarly, the 22 bilateral trade relations affected by Greece's exit from the sample in 2000 do not seem to have any tangible effect. The 40 trade relations affected by Spain and Finland's exit from the sample after 2004 however produce the slump towards a zero RTA effect: the agreements still contributing to the average RTA estimate between 2005 and 2014 are dominated by Mexico with its 36 bilateral partners in the EU/EEA, NAFTA, and the 2005 agreement with Japan.

Returning to Figure 5 the three PPML specifications also produce very different results for the CU effect: for PPML-1 this is indicated to rise from zero to .2 (22%), for PPML-2 it is always negative and statistically insignificant (no CI provided for ease of illustration), whereas for CCE-PPML it drops from around .25 (28%) on introduction of the Euro/ECU to close to zero in 2014. There are no discernable patterns within these sample years.

Overall, the i in Grav$_{it}$y has large implications, not only on its own, but also in combination with the t in Grav$_{it}$y. Results can be drastically different when alternative time windows are used, given that different country-pairs contribute to the identification of the trade policy effect.

5 Concluding remarks

We have shown that the effects of international agreements on trade are considerably heterogeneous across countries, with direct implications for time heterogeneity across different time windows of observation. Our findings suggest that estimates based on pooled estimators are largely uninformative. Fortunately, current research is now moving towards the examination of heterogenous effects of international agreements, allowing policymakers to consider and study those which have had the greatest impact. As argued throughout the paper, adoption of our simple empirical approach would allow for even greater methodological advances, notably in the flexible modelling of spatial dependence and the estimation of country-specific factors. The CCE-PPML estimator is straightforward to apply using the standard `ppml` Stata command by Santos Silva and Tenreyro (2006) in combination with a simple loop for each country pair. Similar benefits from adoption can be expected in other fields (e.g. international finance or international migration) where gravity models are traditionally used.

References

[1] **Anderson, James E., and Eric Van Wincoop. 2003.** 'Gravity with gravitas: a solution to the border puzzle.' *American Economic Review* 93(1): 170-192.

[2] **Andrews, Donald W. K. 2005.** 'Cross-section regression with common shocks.' *Econometrica* 73(5): 1551-1585.

[3] **Bai, Jushan. 2009.** 'Panel data models with interactive fixed effects.' *Econometrica* 77(4): 1229-1279.

[4] **Baier, Scott, and Jeffrey Bergstrand. 2007.** 'Do free trade agreements actually increase members' international yrade?' *Journal of International Economics* 71: 72-95.

[5] **Baier, Scott, Jeffrey Bergstrand and Matthew Clance. 2018.** 'Heterogeneous effects of economic integration agreements.' *Journal of Development Economics*, Vol. 135: 587-608.

[6] **Baier, Scott, Yoto V. Yotov, and Thomas Zylkin. 2019.** 'On the widely differing effects of free trade agreements: Lessons from twenty years of integration.' *Journal of International Economics*, Vol. 116: 206-226.

[7] **Baldwin, Richard E., Virginia Di Nino, Lionel Fontagné, Roberto A. De Santis, and Daria Taglioni. 2008.** 'Study on the impact of the Euro on trade and foreign direct investment.' European Economic and Monetary Union Working Paper 321.

[8] **Barbieri, Katherine, and Omar M. G. Keshk. 2017.** Correlates of war project trade data set (1870-2014). http://correlatesofwar.org.

[9] **Bas, Maria, Thierry Mayer, and Mathias Thoenig. 2017.** 'From micro to macro: Demand, supply, and heterogeneity in the trade elasticity.' *Journal of International Economics* 108: 1-19.

[10] **Behrens, Kristian, Cem Ertur, and Wilfried Koch. 2012.** 'Dual gravity: Using spatial econometrics to control for multilateral resistance.' *Journal of Applied Econometrics* 27(5): 773-794.

[11] **Beverelli, Cosimo, Alexander Keck, Mario Larch, and Yoto Yotov. 2018.** 'Institutions, trade and development: a quantitative analysis.' CESifo Working Paper Series No. 6920.

[12] **Boneva, Lena, and Oliver Linton. 2017.** 'A discrete-choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance.' *Journal of Applied Econometrics* 32(7): 1226-1243.

[13] **Camarero, Mariam, Estrella Gomez and Cecilio Tamarit. 2013.** 'EMU and trade revisited: Long-run evidence using gravity equations.' *The World Economy*, 36(9): 1146-1164.

[14] **Cameron, A. Colin, and Pravin Trivedi. 1998.** *Regression Analysis of Count Data*. Cambridge University Press.

[15] **Campello, Murillo, Antonio F. Galvao, and Ted, Juhl. 2018.** 'Testing for slope heterogeneity bias in panel data models.' *Journal of Business & Economic Statistics*, forthcoming.

[16] **Chen, Natalie, and Dennis Novy. 2018.** 'Currency unions, trade, and heterogeneity.' CEPR Discussion Papers No. 12954.

[17] **Chernozhukov, Victor, Iván Fernández-Val, Jinyong Hahn, and Whitney Newey. 2013.)** 'Average and quantile effects in nonseparable panel models.' *Econometrica* 81(2): 535-580.

[18] **Chudik, Alexander, and M. Hashem Pesaran. 2015a.)** 'Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors.' *Journal of Econometrics* 188(2): 393-420.

[19] **Chudik, Alexander, and M. Hashem Pesaran. 2015b.** 'Large panel data models with cross-sectional dependence: A survey.' In: Badi H. Baltagi (ed) *The Oxford Handbook on Panel Data*. Oxford University Press, Ch. 1.

[20] **Chudik, Alexander, M. Hashem Pesaran, and Elisa Tosetti. 2011.** 'Weak and strong cross section dependence and estimation of large panels.' *Econometrics Journal* 14 (1), C45-C90.

[21] **Eberhardt, Markus, Christian Helmers and Hubert Strauss. 2013.** 'Do spillovers matter when estimating private returns to R&D?' *Review of Economics and Statistics* 95(2), pp. 436-448.

[22] **Eberhardt, Markus and Andrea F. Presbitero. 2015.** 'Public debt and growth: Heterogeneity and non-linearity.' *Journal of International Economics* 97(1), pp. 45-58.

[23] **Eberhardt, Markus and Francis Teal. 2013a.** 'Structural change and cross-country growth empirics.' *The World Bank Economic Review* 27(2), pp. 229-271.

[24] **Eberhardt, Markus and Francis Teal. 2013b.** 'No mangoes in the tundra: Spatial heterogeneity in agricultural productivity analysis.' *Oxford Bulletin of Economics and Statistics*, 75(6): 914-939.

[25] **Fally, Thibault. 2015.** 'Structural gravity and fixed effects.' *Journal of International Economics*, 97: 76-85.

[26] **Fouquin, Michel, and Jules Hugot. 2016.** 'Two centuries of bilateral trade and gravity data: 1827-2014.' CEPII Working Paper No. 2016-14.

[27] **Hamilton, Lawrence C. 1992.** 'How Robust is Robust Regression?' *Stata Technical Bulletin* 1(2): 21-25.

[28] **Head, Keith, and Thierry Mayer. 2014.** 'Gravity equations: Workhorse, toolkit, and codebook.' In: Gita Gopinath, Elhanan Helpman, Kenneth Rogoff (eds) *Handbook of International Economics* Volume 4, Chapter 3.

[29] **Juhl, Ted, and Oleksandr Lugovsky. 2014.** 'A test for slope heterogeneity in fixed effects models.' *Econometric Reviews* 33(8): 906-935.

[30] **Kapetanios, George, M. Hashem Pesaran, and Takashi Yamagata. 2011.** 'Panels with nonstationary multifactor error structures.' *Journal of Econometrics* 160: 326-348.

[31] **Kapetanios, George, Camilla Mastromarco, Laura Serlenga, and Yongcheol Shin. 2017.** 'Modelling in the presence of cross-sectional error dependence.' In: Matyas Laszlo (ed.) *The Econometrics of Multi-dimensional Panels*. Springer, pp. 291-322.

[32] **Kelejian, Harry, George S. Tavlas, and Pavlos Petroulas. 2012.** 'In the neighborhood: The trade effects of the Euro in a spatial framework.' *Regional Science and Urban Economics* 42(1/2): 314-322.

[33] **Klasing, Mariko, Petros Milionis, and Robert Zymek. 2015.** 'Gravity across space and time.' Edinburgh School of Economics, University of Edinburgh, Discussion Paper Series No. 265.

[34] **Koch, Wilfried, and James P. LeSage. 2015.** 'Latent multilateral trade resistance indices: theory and evidence' *Scottish Journal of Political Economy* 62(3): 264-290.

[35] **Mastromarco, Camilla, Laura Serlenga, and Yongcheol Shin. 2016.** 'Multilateral resistance and the Euro effects on trade flows.' In: Roberto Patuelli and Giuseppe (eds) *Spatial Econometric Interaction Modelling*, pp. 253-278. Springer.

[36] **Novy, Dennis. 2013.** 'International trade without CES: Estimating translog gravity.' *Journal of International Economics* 89(2): 271-282.

[37] **Olivero, Maria Pia, and Yoto V. Yotov. 2012.** 'Dynamic gravity: endogenous country size and asset accumulation.' *Canadian Journal of Economics* 45(1): 64-92.

[38] **Pesaran, M. Hashem. 2006.** 'Estimation and inference in large heterogeneous panels with a multifactor error structure.' *Econometrica*, 74(4): 67-1012.

[39] **Pesaran, M. Hashem, and Ron Smith. 1995.** 'Estimating long-run relationships from dynamic heterogeneous panels.' *Journal of Econometrics* 68(1): 79-113.

[40] **Pesaran, M. Hashem and Elisa Tosetti. 2011.** 'Large panels with common factors and spatial correlation.' *Journal of Econometrics*, 161(2), pp. 182-202.

[41] **Piermartini, Roberta, and Yoto Yotov. 2016.** 'Estimating trade policy effects with structural gravity.' In: Yotov, Yoto, Roberta Piermartini, José-Antonio Monteiro, and Mario Larch (eds) *An advanced guide to trade policy analysis: The structural gravity model*. Geneva: World Trade Organization.

[42] **Santos Silva, João M. C., and Silvana Tenreyro. 2006.** 'The log of gravity.' *Review of Economics and Statistics* 88(4): 641-658.

[43] **Santos Silva, João M. C., and Silvana Tenreyro. 2010.** 'Currency Unions in Prospect and Retrospect.' *Annual Review of Economics* 2(1): 51-74.

[44] **Serlenga, Laura, and Yongcheol Shin. 2007.** 'Gravity models of Intra-EU trade: Application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors.' *Journal of Applied Econometrics* 22: 361-381.

[45] **Stock, James H., and Mark W.Watson. 2002.** 'Forecasting using principal components from a large number of predictors.' *Journal of the American Statistical Association* 97(460):1167-1179.

[46] **Westerlund, Joakim. 2018.** 'On estimation and inference in heterogeneous panel regressions with interactive effects.' *Journal of Time Series Analysis*, forthcoming.

Tables and Figures

Table 1: RTA and CU – Sample Policy Evolution

RTA					CU				
Policy Regime	Start Date	Years in regime	Pairs w/ RTA=1		Policy Regime	Start Date	Years in regime	Pairs w/ COMCUR=1	
I	1960Q1	13	54	14.2%	I	1960Q1	19	2	0.5%
II	1973Q1	1	132	34.7%					
III	1974Q1	7	146	38.4%					
					II	1979Q1	20	0	0.0%
IV	1981Q1	5	168	44.2%					
V	1986Q1	3	208	54.7%					
VI	1989Q1	5	210	55.3%					
VII	1994Q1	4	216	56.8%					
VIII	1998Q1	4	242	63.7%					
					III	1999Q1	2	72	18.9%
					IV	2001Q1	14	90	23.7%
IX	2002Q1	3	246	64.7%					
X	2005Q1	5	250	65.8%					
XI	2010Q1	5	256	67.4%					
	2014Q4	end				2014Q4	end		

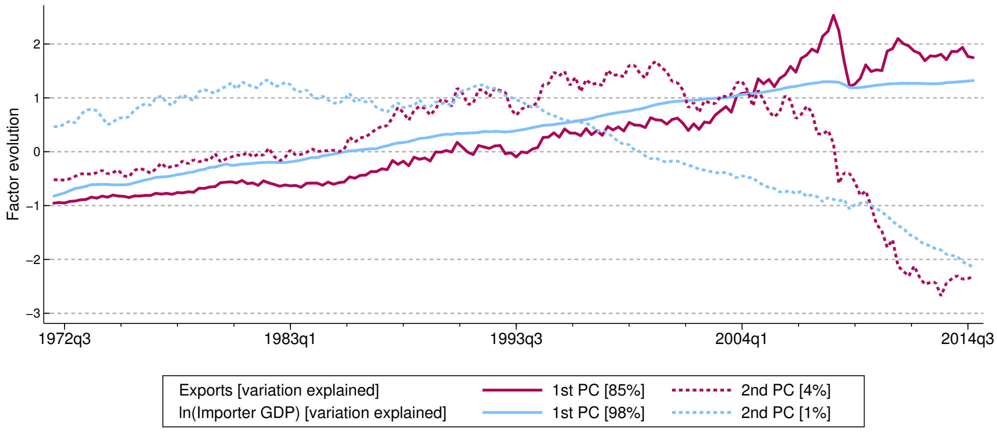
Notes: This table identifies all the changes to the RTA (left panel) and CU (right panel) dummies employed in the empirical analysis for our sample of 380 country pairs: in 1960 54 pairs or 14% of the sample country pairs had RTAs in place, a policy 'regime' which stayed unchanged until 1973 when 78 additional country pairs entered into RTAs, such that the set of country pairs with RTAs increased to 132 (35%). With regards to RTAs we observe a total of ten changes (eleven regimes) in the set of countries with this policy instrument in place, whereas for CU we see a mere three changes, which are furthermore comparatively limited in scope (dissolution of GBR-IRL currency union in 1979, two waves of ECU/Euro area creation and expansion in 1999 and 2001). Between 14 and 67% of the sample country pairs had RTAs, between 0 and 24% CU arrangements. The identification of the RTA and CU effects in the full time series and rolling window analyses are off these 'regime' changes – Figure 2 provides more details.

Table 2: Full Time Series Gravity Regressions

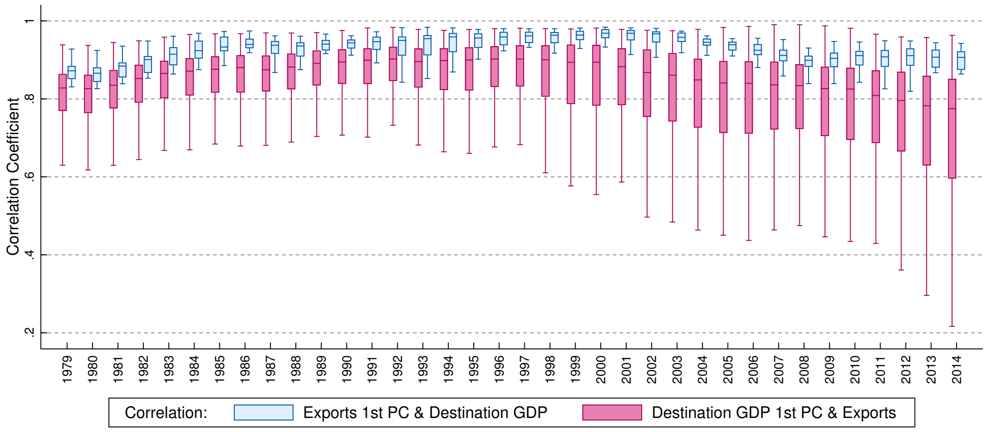
Specification	Pooled PPML w/ pairwise FE			Pooled PPML w/ pairwise & directional-time FE			Heterogeneous CCE-PPML		
	[1]	[2]	[3]	[4]	[5]	[6]		[7]	[8]
Data Frequency	Quarterly	Annual (Q3)	Tri-annual	Quarterly	Annual (Q3)	Tri-annual	pairs	Quarterly	Annual (Q3)
ln(Origin GDP)	0.847	0.884	0.866				380	0.539	0.565
	[0.242]***	[0.254]***	[0.241]***					[0.128]***	[0.136]***
ln(Destination GDP)	0.864	0.844	0.884				380	1.638	1.685
	[0.194]***	[0.207]***	[0.213]***					[0.127]***	[0.137]***
RTA	0.424	0.438	0.468	0.292	0.293	0.299	202	0.208	0.189
	[0.129]***	[0.129]***	[0.138]***	[0.010]***	[0.079]***	[0.082]***		[0.031]***	[0.031]***
in %age terms	52.8%	54.9%	59.7%	33.9%	34.0%	34.8%		23.1%	20.9%
CU	0.244	0.250	0.234	-0.028	-0.038	-0.026	92	-0.117	-0.141
	[0.053]***	[0.057]***	[0.057]***	[0.009]***	[0.052]	[0.059]		[0.073]	[0.081]*
in %age terms	27.6%	28.4%	26.3%	-2.8%	-3.7%	-2.5%		-11.0%	-13.1%
Time FE	Yes	Yes	Yes	Yes	Yes	Yes		No	No
Pair FE	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes
Directional-time FE	No	No	No	Yes	Yes	Yes		No	No
CA augmentation	No	No	No	No	No	No		Yes	Yes
Observations	83,420	20,846	6,840	83,412	20,843	6,840		82,994	20,734
Trade flow pairs	380	380	380	380	380	380		380	380
Average T	219.5	54.9	18.0	219.5	54.9	18.0		218.4	54.6

Notes: We present PPML estimation results for the full time series from 1960Q1 to 2014Q4, adopting alternative data frequencies: the reported 'Quarterly' data, 'Annual' data using the third quarter in each year, 'Tri-Annual' data. Fixed effects and other augmentations of the estimation equation are indicated in a lower panel, along with the sample characteristics; 'Average T' reports the average pair-specific time series in the sample. The dependent variable in each model is exports. In the models in [1]-[6] the standard errors are clustered at the country-pair level. In the PPML-CCE model we report robust mean average across the 380 country-pair regressions (Hamilton, 1992; Pesaran and Smith, 1995); here the trade policy variables are only identified in those country pairs for which any policy change occurred over time. In pair-specific regressions where this is not the case the related trade policy variable drops out – hence the lower pair count as indicated. Alongside the gravity model estimates for the RTA and CU dummies we report the trade volue effects (in percent).

Figure 1: Correlation Analysis — Common Factors and Observables



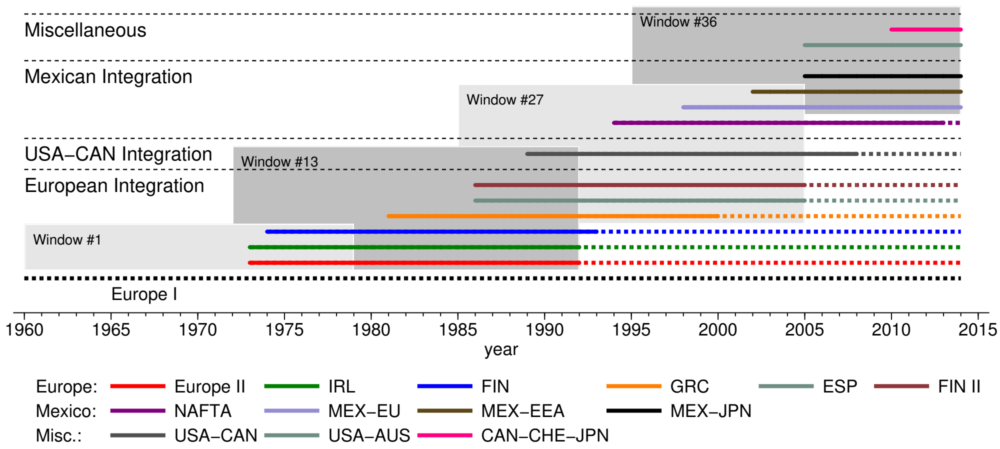
(a) Evolution of Common Factors



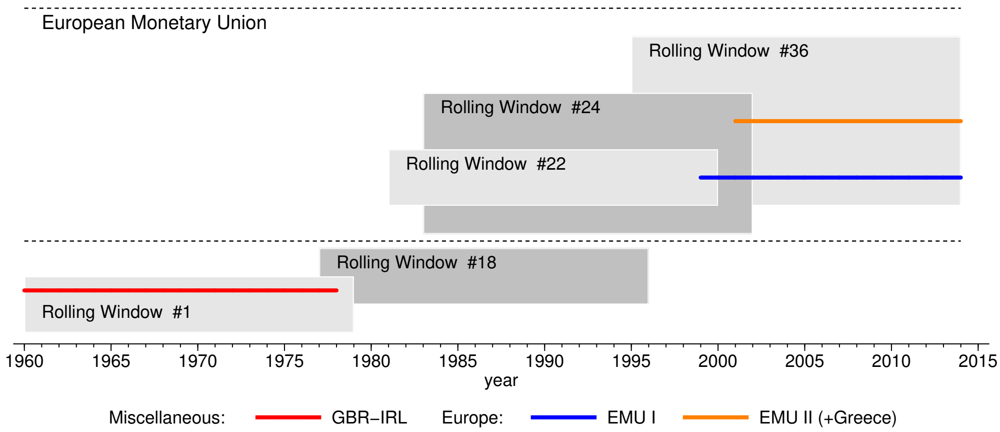
(b) Correlation between Common Factors and Observables

Notes: Panel (a) plots the first and second principal components for exports and destination GDP, which account for 89% and 99% of variation in these variable series, respectively. Panel (b) presents two series of box plots of correlation coefficients (by end year): (i) the correlation coefficient between the first principal component of exports and the observed value for destination GDP (in logs); (ii) the correlation coefficient between the first principal component of destination GDP (in logs) and the observed value for exports (i.e. the imports into the destination country). For the end years until 1998 there are 362 correlation coefficients making up each box plot, from 1999 onwards there are 380. The box represents the interquartile range, the whiskers extend to the top and bottom percentile.

Figure 2: Identifying Policy Effects in the Rolling Window Analysis



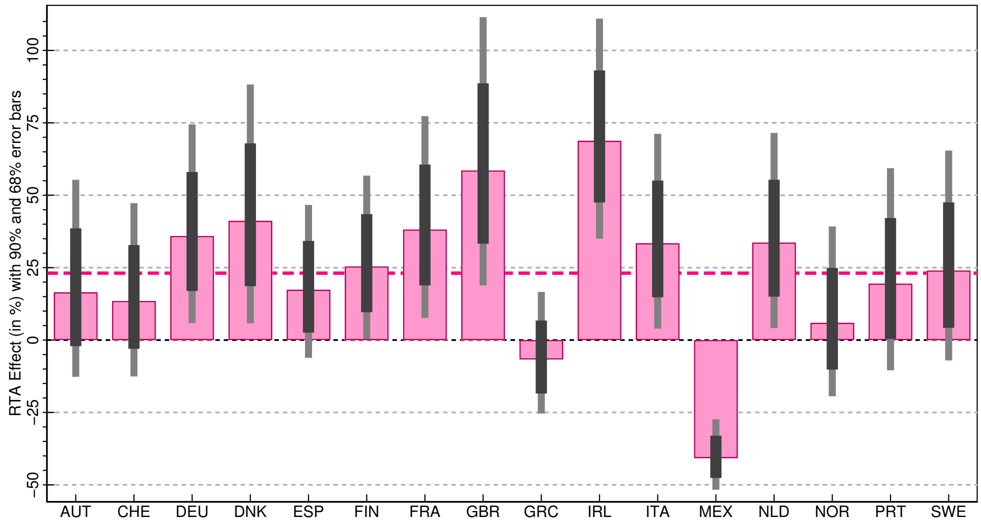
(a) RTAs



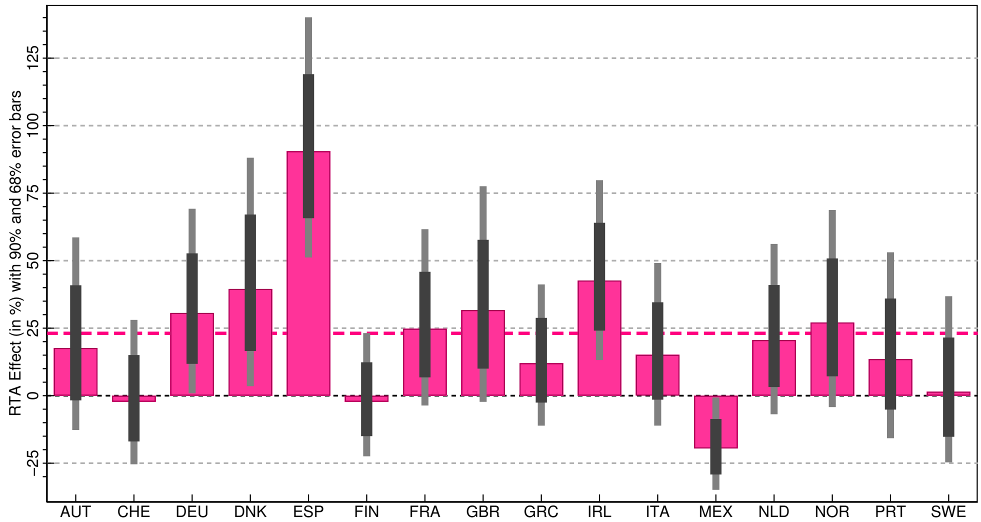
(b) CU

Notes: These plots help to summarize the identification of trade policy effects in the rolling window analysis. They are specific to the sample of countries applied in this study and presented separately for RTAs and CU agreements in Panels (a) and (b) ,respectively. The crucial point for the rolling window analysis is that a trade policy effect is only identified for those country pairs which *entered into an agreement* during the time period covered by the specific window. Each plot highlights *in solid lines* the first 20 years for each new wave of trade agreements alongside a number of rolling windows (20 years or 80 quarters in length) over the 1960-2014 time horizon. After the first 19 years the specific agreement or set of agreements does no longer contribute to the identification of the policy effect since it *does not change* over the rolling window of analysis – we therefore highlight the continuation of trade policy beyond 20 years *with dashed lines*. For instance, rolling window #1 sees a group of European countries (Europe II, red line), with Ireland (IRL, green) and Finland (FIN, blue) joining later, towards the end of this 20-year window. From Window #14 onwards the Europe II, IRL and FIN agreements no longer contribute to identifying the RTA effect: it is identified by other, later agreements. Note that the first wave of European integration (Europe I) does not help identify *any* RTA effect since this was already in place by 1960Q1.

Figure 3: Country-Specific Trade Policy Effects for RTAs (CCE-PPML)



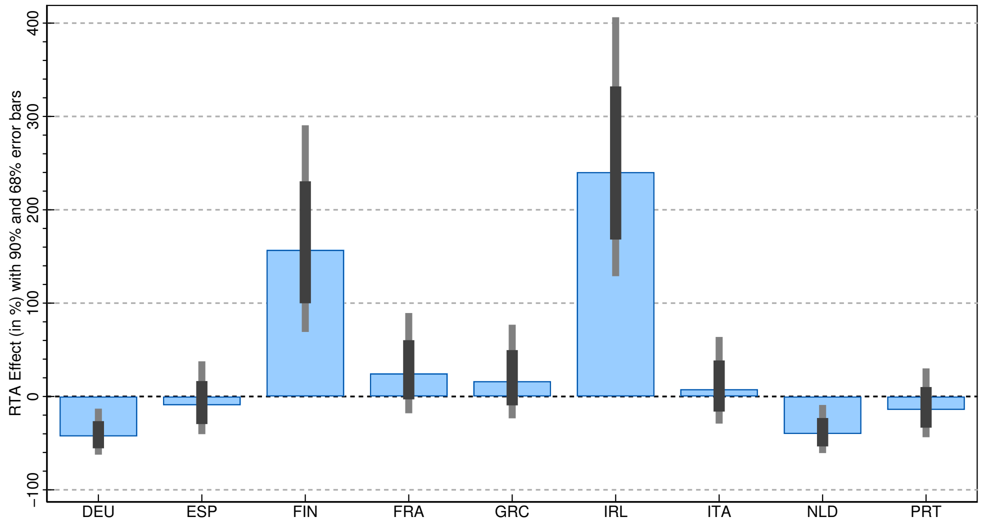
(a) RTA Effects from the perspective of exports



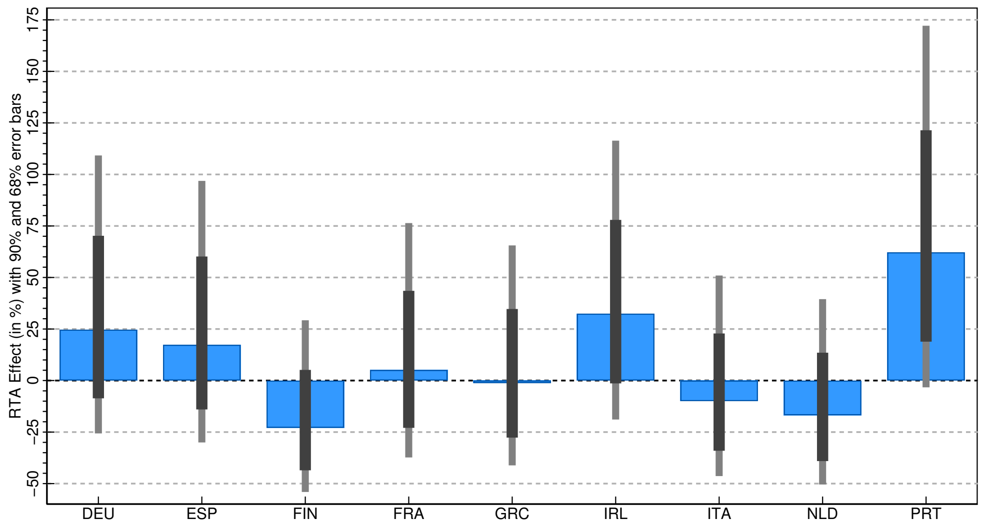
(b) RTA Effects from the perspective of imports

Notes: These plots present country means for the RTA effect from the perspective of exports and imports in panels (a) and (b), respectively. The underlying CCE-PPML estimates at the pair-level are based on the model presented in Table 2 Column 7 — the robust mean of the RTA estimate is indicate with a dashed horizontal line. Some countries are omitted here since their country estimate (as exporter or importer) would only be based on a small number of pair-level estimates (see Figure 2): AUS (1 RTA with USA), CAN (4 RTAs with USA, CHE/JPN, USA, and MEX via NAFTA), JPN (2 RTAs with CAN/CHE and MEX), and USA (3 RTAs with AUS, CAN, and MEX via NAFTA). The other country results presented are Mean Group estimates computed from between 9 and 18 country-pair results. In the light of these small sample sizes we present two sets of confidence intervals for 90% (light grey bars) and 68% (dark grey bars) — these are not symmetric as the estimation results are exponentiated. Estimating weighted averages (based on robust regression weights) yields qualitatively identical results.

Figure 4: Country-Specific Trade Policy Effects for EMU (CCE-PPML)



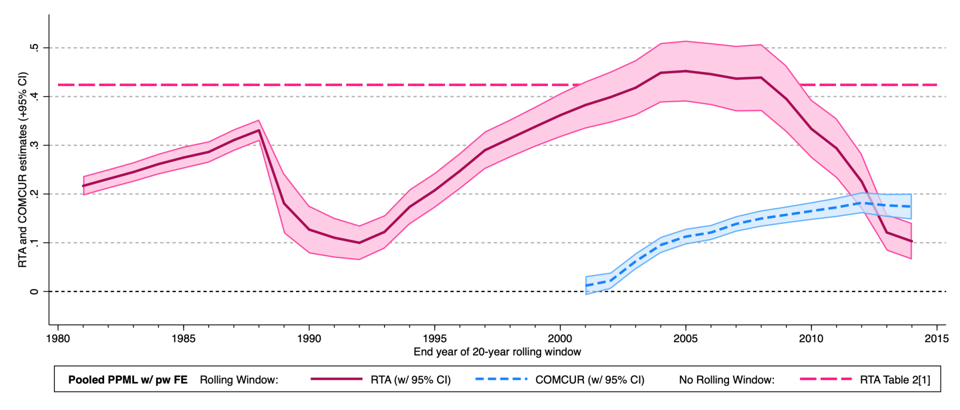
(a) CU Effects from the perspective of exports



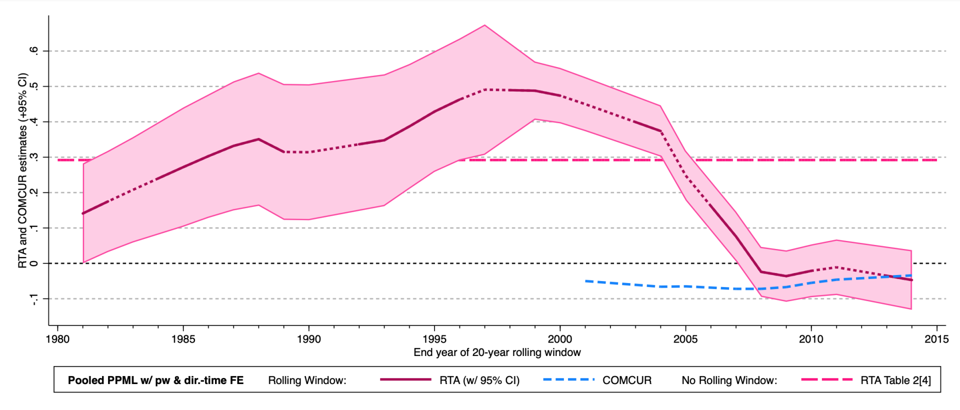
(b) CU Effects from the perspective of imports

Notes: These plots present country means for the COMCUR/Euro effect from the perspective of exports and imports in panels (a) and (b), respectively. The underlying CCE-PPML estimates at the pair-level are based on the model presented in Table 2 Column 7. Some countries are omitted here since their country estimate (as exporter or importer) would only be based on a small number of pair-level estimates. The other country results presented are Mean Group estimates computed from between 9 and 10 (for IRL) country-pair results. In the light of these small sample sizes we present two sets of confidence intervals for 90% (light grey bars) and 68% (dark grey bars) — these are not symmetric as the estimation results are exponentiated. Estimating weighted averages (based on robust regression weights) yields qualitatively identical results.

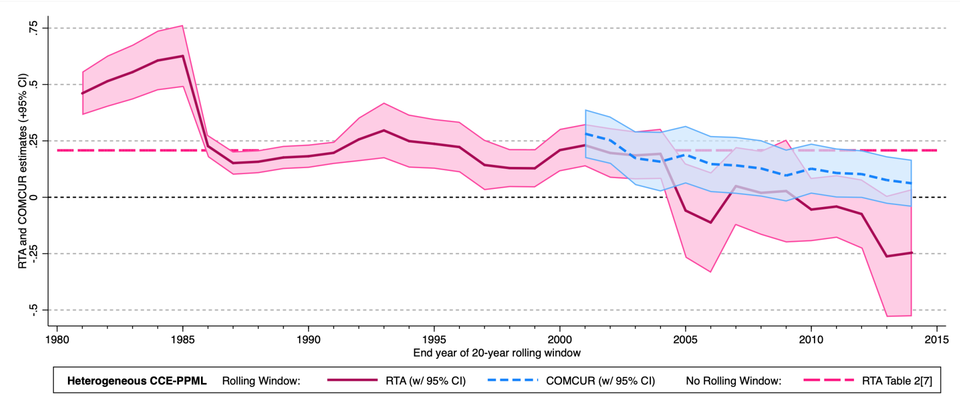
Figure 5: Rolling Regression Results



(a) Pooled PPML with Pairwise FE



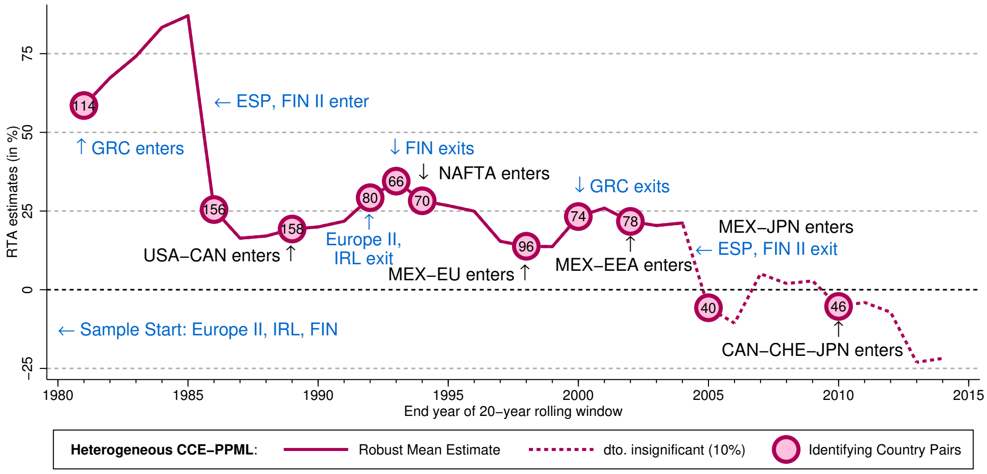
(b) Pooled PPML with Pairwise FE and Directional-Time FE



(c) Heterogeneous CCE-PPML

Notes: The graphs present estimates for the RTA and CU effects from pooled PPML and heterogeneous PPML-CCE gravity model regressions. The window length is 20 years (80 quarters) and we shift the window by one year at a time; the x-axis reports the end-year of each rolling window estimate, e.g. 1981 implies 1981Q4 is the final observation. The COMCUR CI in (b) is omitted for ease of illustration (it *always* contains zero). The short-dashed RTA line in (b) indicates interpolated estimates: for these windows the Pooled PPML model did not converge.

Figure 6: Rolling Regression Results (CCE-PPML) and Policy Regimes



Notes: The graph presents the robust mean estimates for RTA from the heterogeneous PPML-CCE gravity regressions. The window length is 20 years (80 quarters) and we shift the window by one year at a time; the x-axis reports the end-year of each rolling window estimate, e.g. 1981 implies 1981Q4 is the final observation. The dashed line indicates when the robust mean estimate is statistically insignificant at the 5% level. Additional text highlights which country-pairs entered and which dropped out of the rolling window regression sample (no RTAs are discontinued; exit always refers to exit from the sample and not the agreement); we use blue text to indicate all RTAs which related to agreements *within* the EU/EEA, and black text for all others (including non-EU/EEA countries' agreements with the EU/EEA).

APPENDIX

A Data Appendix

Table A-1: Descriptive Statistics

		Obs	Pairs	Mean	Median	SD	Min	Max
Export value	exports	83,420	380	1.16E+09	1.38E+08	4.3E+09	0	9.67E+10
GDP	lngdp	83,420	380	26.92	26.67	1.30	23.77	30.43
RTA	rta	83,420	380	0.45	0	n/a	0	1
CU	comcur	83,420	380	0.07	0	n/a	0	1

Notes: The table details standard descriptive statistics for the regression sample. SD refers to the standard deviation. Destination and Origin GDP are minimally different and hence not reported separately.

Table A-2: Sample Makeup

ISO code	Country	Obs	Missing flows
AUS	Australia	5,353	
AUT	Austria	5,353	
CAN	Canada	5,353	
CHE	Switzerland	5,353	
DEU	Germany	5,353	
DNK	Denmark	5,353	
ESP	Spain	5,353	
FIN	Finland	5,353	
FRA	France	5,353	
GBR	United Kingdom	5,353	
GRC	Greece	5,353	
IRL	Ireland	5,353	
ITA	Italy	5,353	
JPN	Japan	5,353	
MEX	Mexico	5,123	1972, 1974Q3/Q4, 1979♯
NLD	Netherlands	5,353	
NOR	Norway	5,353	
PRT	Portugal	5,353	
SWE	Sweden	5,353	
USA	United States	5,353	

Notes: The table lists the sample countries for which we study bilateral trade flows over the 1960-2014 time horizon. ♯ The Mexican data lacks export flows to all countries with the exception of Germany in the years/quarters indicated.