Untangling Fixed Effects and Constant Regressors

Franc Klaassen
University of Amsterdam and Tinbergen Institute
&
Rutger Teulings *
CPB Netherlands Bureau for Economic Policy Analysis
and University of Amsterdam

July 29, 2018

Abstract

Fixed effects (FE) in panel data models overlap each other and prohibit identification of the impacts of "constant" regressors. Think of regressors that are constant across countries in a country-time panel with time FE. There is identification, however, if the normalized FE are zero. This constraint is more plausible for normalizations that shrink the FE of focus. We thus introduce "untangling normalization", which orthogonalizes the FE to each other and to constant regressors. Untangled FE do not overlap, easing interpretation. In a gravity model for bilateral exports to the US the constant regressors US GDP, world GDP, and US effective exchange rate explain 98% of the time FE, making the FE redundant. We thus achieve identification, even though that is commonly considered impossible.

Key words: fixed effects, gravity model, identification, multicollinearity, normalization, orthogonalization.
JEL classification: C18; C23; F14.

---

*Corresponding author. Address: Bezuidenhoutseweg 30, 2594 AV The Hague, The Netherlands; tel. +31-6-46988073; e-mail R.M.Teulings@cpb.nl
An earlier version has appeared as Tinbergen Institute Discussion Paper 15-137/VI.
We would like to thank Peter Boswijk, Maurice Bun, Noud van Giersbergen, Artūras Juodis, Frank Kleibergen, Maurice Koster, Jan Magnus, Christopher Taber, Andrey Vasnev, and seminar participants for their valuable comments to help improve our paper.

1 Introduction

Panel data models often include fixed effects (FE) to control for unobserved heterogeneity. Estimates of the FE themselves are typically ignored. Still, they contain valuable information about determinants of the dependent variable that are omitted from the model.

The goal of the paper is to extract this information. We do so in two ways, that is, by easing the interpretation of the FE, and by facilitating the identification of the true values of the impacts of "constant" regressors, which are regressors that are invariant in a dimension of the data but vary in another dimension.

The idea is as follows. To avoid perfect multicollinearity, some normalization is needed. Normalization does not identify the true values. However, we derive a testable constraint on normalized FE that yields identification. We then introduce "untangling normalization", which makes the FE orthogonal to each other and to constant regressors. The next paragraphs illustrate how this helps interpretation and can make the identifying constraint more reasonable.

First, consider the problem of interpreting estimated FE. For simplicity, take a country-time panel model with $\alpha + \alpha_i$, where α is the intercept and α_i is the FE for country i. Perfect multicollinearity is commonly prevented by leaving out some FE. We call this zero normalization, denoted by a 0 superscript, and an example is $\alpha^0 = 0$. Now the α_i^0 capture the effect of country i plus the overall intercept, so they overlap. This blurs the signal from omitted determinants in the estimated FE and thus hampers interpretation.

Our first contribution is to resolve overlap by untangling normalization. This here means setting the mean of the country FE to zero, so that the untangled constant α^u captures the overall intercept, and the untangled country FE α_i^u is the country deviation from the overall intercept. There is no overlap anymore, easing interpretation. This specific example is not new; see Suits (1984). But we use a much richer panel setting, with country FE, time FE, country-specific trends, and regressors that are constant across countries or time, and untangling can be generalized further, for example, to three-dimensional panels. We also address identification and estimation.

The second problem of extracting information from estimated FE concerns identification. For example, from the sum $\alpha_i + v_i'\nu$, where v_i is the constant regressor vector with impact ν, one cannot infer ν. Perfect multicollinearity is again avoided by imposing a normalization. A popular example is to normalize

$\nu^0 = 0$, that is, leave out the constant regressors. Whatever normalization, one ends up with a pseudo-true value. The true value ν, which is the value of interest, is not identified. This is a notorious problem in the literature.

For our second contribution, consider the normalized FE. If they are zero, then the true value ν is identified, as we will show. So the normalization by itself does not yield identification, of course, but the constraint does. Still, the normalization can matter for the validity of the constraint. For example, under normalization $\nu^0 = 0$, the constraint $\alpha_i^0 = 0$ for all i says that no country variation exists, which is unlikely. In contrast, untangling normalization exploits the information in the constant regressors v_i by defining α_i^u as the remainder of a projection of the country FE on the constant regressors. This cleans and shrinks the normalized FE, making the identifying constraint $\alpha_i^u = 0$ more realistic. Whether it holds in a specific application is an empirical question, and we present two Wald tests for it. If the constraint holds, the true value is identified, a result commonly considered beyond reach. Otherwise, identification fails, but by controlling for the constant regressors the estimated FE still better signal what the model misses.

Our orthogonalization of FE regarding constant regressors looks like that in existing approaches, such as the random effects (RE), Hausman and Taylor (1981), Plümper and Troeger (2007), Pesaran and Zhou (2016), and Honoré and Kesina (2017) estimators. The key difference is that in all existing approaches the orthogonality relations are identifying restrictions, whereas our method imposes them as just normalizations and we use another restriction for identification, which we can test. That is, other approaches at some point assume RE-types of restrictions, assuming some regressors to be uncorrelated with the effects, with the advantage of increased estimation efficiency if the restrictions hold. We work in a fully FE setting, allowing all regressors to correlate freely with the effects. This is typically considered an important advantage of FE over RE.

There are also practical advantages of our approach. Computation of untangled estimates is a linear transformation of zero-normalized estimates, so there is no extra estimation step. Untangling also conveniently generalizes to other types of FE and provides insights into omitted regressors.

We apply untangling normalization to a gravity model for exports from 17 OECD countries to the US from 1979-2011; see Anderson and van Wincoop (2003) for gravity theory. The estimated untangled time FE are indeed easy to interpret. They reveal that the constant regressor US GDP is the key driver of the time FE. Controlling for it unlocks the importance of the US effective exchange rate, typically omitted from the gravity model. The constant regressors explain 98% of the time effects, and the remaining untangled time FE are insignificant and can be left out. We thus identify the true values of the impacts of the constant regressors, despite the general notion that this is beyond reach. Furthermore, the relevance of the US effective exchange rate has stimulated Klaassen and Teulings (2017) to extend the gravity theory, and they show that the effective exchange rate is a building block of the US multilateral import resistance. The latter is unobserved, which is an important issue in the gravity literature, but untangling identifies it (up to a constant and trend).

The paper is organized as follows. In Section 2 we introduce the model and discuss identification. Section 3 sets out untangling normalization, discusses its advantages, and compares it to the literature. In Section 4 we discuss the applicability of our method and two ways to estimate untangled parameters, while Section 5 develops two tests for identification. In Section 6 we apply untangling normalization to the gravity model. Section 7 concludes.

2 Model specification and identification

Throughout the paper we consider a two-dimensional balanced panel model with dimensions i and t, representing country and time, say. There are N countries and T time periods. All vectors are column vectors unless stated otherwise.

2.1 Model

The dependent variable y_{it} is modeled by

$$y_{it} = \alpha + \alpha_i + \tau \cdot t + \tau_i \cdot t + \theta_t + v_i' \nu + w_t' \omega + x_{it}' \beta + \varepsilon_{it}, \tag{1}$$

where for the sake of exposition all fixed effects (FE) and constant regressors v_i and w_t are mentioned.

The model has three FE-families, each targeting a specific variation of the data. The α-family targets the level variation across countries. It has a homogeneous type, α, and a heterogeneous type, α_i. The τ-family targets the linear trends across countries, with a homogeneous type, τ, and a heterogeneous type, τ_i. The θ-family targets the time variation, with only a heterogeneous type, θ_t.

Model (1) also has i-, t-, and it-regressors. The vector v_i of length K_v depicts all variables that vary over countries but are constant over time. Similarly, w_t of length K_w captures all variables that vary over time but are constant over countries. Variables that vary over both dimensions are in x_{it} of length K_x.

Finally, ε_{it} is the error term. Our untangling approach does not depend on or affect the error term, so the assumptions on it are those that the user imposes.

It is convenient to write (1) in matrix form, where the time series of the countries are stacked. All K_d deterministic variables are contained in the matrix D and the FE parameters in vector $\delta = [\alpha, \alpha_1, \ldots, \alpha_N, \tau, \tau_1, \ldots, \tau_N, \theta_1, \ldots, \theta_T]'$. The constant regressors are stacked into the matrix Z, such that row it becomes $[v_i', w_t']$, consisting of $K_z = K_v + K_w$ columns, and its corresponding parameter vector is $\gamma = [\nu', \omega']'$. All it-regressors are stacked in X. Hence

$$y = D\delta + Z\gamma + X\beta + \varepsilon, \tag{2}$$

where y and ε stack all y_{it} and ε_{it}, respectively. We assume that the columns in $[D, Z, X]$ are linearly independent, except for the dependencies in D and between D and Z set out in the next section (Z itself has full column rank).

2.2 Multicollinearity and the need for normalization

There is (perfect) multicollinearity in $[D, Z]$ for two reasons. The first is within D. For example, the vector of ones in D is the sum of the N country dummies in D. In general, D has column rank $K_d - m_d$, where m_d is the degree of multicollinearity, the number of dependent columns. In model (2) $m_d = 4$, that is, one due to the α-family, one due to the τ-family, and two because θ_t is combined with α and τ.

The second source of multicollinearity is that the v_i-columns in Z are linear combinations of the country dummies in D, and similarly for the w_t-columns regarding the time dummies. This adds $m_z = K_z$ dependencies.

In total, the column rank of $[D, Z]$ is $K_d + K_z - m_d - m_z$. So we need $m_d + m_z$ normalizations to estimate the model. They do not influence estimation of β. The focus in this paper is thus on $D\delta + Z\gamma$, in particular on how to distribute this sum over the fixed effects and the constant regressors.

2.3 Identifying the impacts of constant regressors

Normalization resolves multicollinearity, but there is still an identification problem. To analyze this, let us focus on v_i; the approach for w_t is similar.

The identification problem is that an observationally equivalent model results by taking some other value ν^* instead of ν, defining $\alpha^* + \alpha_i^*$ such that

$$\alpha + \alpha_i + v_i'\nu = \alpha^* + \alpha_i^* + v_i'\nu^*, \tag{3}$$

and then substituting (3) into (1). In particular, using some normalization one can estimate the associated pseudo-true values α^*, α_i^*, and ν^*, but one cannot infer estimates of the true values α, α_i, and ν from them. This is a well-known and unsolved problem.

As a potential solution, consider the constraint $\alpha_i^* = 0$ for all i. Under this null hypothesis, (3) becomes

$$\alpha + \alpha_i = \alpha^* + v_i' \left(\nu^* - \nu\right), \tag{4}$$

so that the unobserved $\alpha + \alpha_i$ is fully driven by the constant regressors. Because there is no exact linear relationship among those regressors, $\nu^* - \nu$ is unique.

The (ceteris paribus) impacts of all variables in v_i on y_{it} in model (1) are given by ν plus the effect through α_i, which is $\nu^* - \nu$. Hence the impacts of the constant regressors are identified and equal to ν^*. Put differently, $\alpha_i^* = 0$ implies that the true value $\nu = \nu^*$ and using (4) then gives that $\alpha + \alpha_i = \alpha^*$ does not depend on i. So we have a testable constraint to tackle the identification problem.

The above holds for any normalization.[1] If the normalization only involves α_i^*, then the null hypothesis constrains the remaining $N - 1 - K_v$ elements of α_i^*. This correctly reflects that under the null the N values of (3) are determined by $1 + K_v$ parameters. However, normalizing $\alpha^* = 0$ and $\nu^* = 0$ implies that the null imposes N constraints, making it less realistic. Put differently, α^* and ν^* can no longer control for the constant term and constant regressors, so that the latter end up in α_i^*; it is better to split off those disturbing influences and thereby shrink α_i^*. In general, for a clean identification analysis one should normalize parameters that will be constrained by the null hypothesis.

The next section introduces a normalization that fulfills this requirement and is particularly convenient for interpretation, and in Section 5 we present tests for the constraint. Section 6.5 illustrates how our approach can indeed yield identification in practice.

[1] We exclude normalizations that make the null hypothesis impossible, for example, normalizations that fix α_i^* at a non-zero value for some i.

3 Untangling normalization

3.1 The idea and advantages of untangling

The idea of untangling normalization is to prevent multicollinearity by making the FE orthogonal to each other and to constant regressors. We can now interpret the FE as deviations from the other FE and the constant regressors; they have been untangled. This section sets out the two aspects of our approach — untangling a FE from other FE and from constant regressors — and describes the advantages regarding interpretation and identification. Untangled parameters have a u superscript.

First, consider α and α_i as an example, so that we need one normalization. Untangling normalization sets the mean of the latter to zero. Now the untangled constant α^u captures the overall level, and the untangled country FE α_i^u is the country deviation from the overall level. So both effects do not interfere with each other and are assigned to separate parameters, in a unique way. Similarly, $\tau^u \cdot t$ controls for the overall trend, $\tau_i^u \cdot t$ is the country deviation from that, and θ_t^u captures the time deviation from the overall level and trend.

Untangling thus eases interpretation. Estimates of, say, α_i^u may also help to find potentially important omitted regressors. These are advantages over other normalizations, such as zero normalization, where the overall level and the country deviations are "tangled" into the FE.

The second aspect of untangling concerns the relation between α_i and the country-specific regressors v_i. We need K_v normalizations. Untangling normalization sets the country FE orthogonal to the variables in v_i. Now the α_i^u capture what is left over after the explanation by v_i, thereby exploiting the information in constant regressors. Likewise, θ_t^u captures what is left over after using w_t.

We can now estimate ν^u. This is a pseudo-true value, reflecting the identification problem discussed in Section 2.3. There we also derived that, if $\alpha_i^u = 0$ for all i, then $\nu^u = \nu$, so that the true value is identified. Moreover, we showed that for a clean identification test the normalization should be on the parameters constrained by the null hypothesis. Untangling normalization fulfills this, because it normalizes α_i^u, not α^u or ν^u. Whether $\alpha_i^u = 0$ holds in a specific application is an empirical question. If it is rejected, there is no identification, but the estimated α_i^u still indicate what the model misses in that particular direction, exemplifying the benefits of untangling regarding interpretation.

3.2 Untangling fixed effects

We first introduce all untangling normalizations for the FE-types in our model.

Common constant: α

Because α is the homogeneous type of the intercept fixed effects, we want α^u to capture the overall intercept in the model, so we do not normalize it.

Country-specific effects: α_i

To untangle the country FE α_i from the common constant α, we normalize their mean to zero, so that the untangled α_i^u capture the country deviations from α^u:

$$\sum_i \alpha_i^u = 0. \tag{5}$$

Common trend: $\tau \cdot t$

Just as for α, τ is the homogeneous type of the trend fixed effects, so that we want τ^u to capture the overall trend in the model and do not normalize it.

Country-specific trends: $\tau_i \cdot t$

Similar to α_i, we normalize the mean of the country-trend FE τ_i to zero, so that the untangled $\tau_i^u \cdot t$ capture the country deviations from the common trend $\tau^u \cdot t$:

$$\sum_i \tau_i^u = 0. \tag{6}$$

Time-specific effects: θ_t

Similar to α_i, we normalize the mean of θ_t to zero, so that the θ_t^u are the time deviations from the overall intercept α^u. In addition, time FE pick up the common trend. Because we already have the latter in the model, we untangle θ_t from $\tau \cdot t$ by orthogonalizing θ_t to the trend. This ensures that θ_t^u is trendless and is the time deviation from the common trend. In formula,

$$\sum_t \theta_t^u = 0 \tag{7}$$

$$\sum_t \theta_t^u \cdot t = 0. \tag{8}$$

3.3 Untangling to exploit constant regressors

We also define the untangling normalization for FE in relation to constant regressors that vary solely over the same dimension.

Country-specific regressors: v_i
To clean country FE from the information in v_i, we make α_i orthogonal to v_i. We get the following K_v normalizations, one for each regressor v_i^k:

$$\sum_i \alpha_i^u v_i^k = 0. \tag{9}$$

Time-specific regressors: w_t
Similarly, we clean the time FE from w_t, resulting in K_w normalizations

$$\sum_t \theta_t^u w_t^k = 0. \tag{10}$$

3.4 Comparison to the literature: normalizations vs. restrictions

The above orthogonality relations concern the untangling of FE regarding each other and regarding the constant regressors. The first group are a generalization of Suits (1984), who uses only (5). The second group, (9) and (10), look like those in existing approaches such as the random effects (RE), Hausman and Taylor (1981), and Plümper and Troeger (2007) estimators, but there are several profound differences.

The key difference is that in all existing approaches the orthogonality relations are identifying restrictions, whereas our method imposes them as just normalizations. That is, other approaches at some point assume RE-types of restrictions, restricting some regressors to be uncorrelated with the effects, with the advantage of increased estimation efficiency. In contrast, we work in a fully FE setting, allowing all regressors to correlate freely with the effects. This is typically considered an important advantage of FE over RE. We now discuss the differences in more detail, focusing on α_i and constant regressor v_i in model (1). If we refer to existing methods, α_i is considered a random variable.

The RE approach imposes moment conditions $\mathbb{E}\left\{x_{it}\alpha_i\right\} = 0$ and $\mathbb{E}\left\{v_i\alpha_i\right\} = 0$. Untangling normalization does not need orthogonality of x_{it} and α_i. Moreover, we do not impose orthogonality regarding v_i as an identifying restriction, but only as a normalization.

Hausman and Taylor (1981) weaken the RE restrictions by imposing them only for known subsets of the it-regressors and constant regressors. Denote those subsets by a superscript 1. Then the moment conditions are $\mathbb{E}\left\{x_{it}^1\alpha_i\right\} = 0$ and $\mathbb{E}\left\{v_i^1\alpha_i\right\} = 0$. The means of x_{it}^1 over time are then used as instruments for the other variables in v_i, which are allowed to correlate with α_i. This requires that the means of x_{it}^1 are sufficiently correlated with those endogenous constant regressors. Finding x_{it}^1 that fulfill both requirements is onerous (Breusch et al. (2011)), and that has motivated Plümper and Troeger (2007) to focus on methods that do not rely on such instruments.

Plümper and Troeger (2007) introduce the fixed effects vector decomposition (FEVD). This allows α_i to freely correlate with the it-regressors but imposes zero correlation with all constant regressors, that is, $\mathbb{E}\left\{v_i\alpha_i\right\} = 0$. So FEVD is a hybrid of FE and RE, as stressed by Greene (2011). A similar combination underlies the estimators by Plümper and Troeger (2011), Pesaran and Zhou (2016), and Honoré and Kesina (2017), in contrast to our fully FE approach.

There are also practical differences between the literature, represented by FEVD, and our approach. First, consider the implementation of FEVD. That entails two steps, where step one computes the within-groups estimator to obtain the associated residuals, and step two regresses their time-averages on the constant regressor v_i. Note that this is FEVD after the correction of the second step by Plümper and Troeger (2011). Breusch et al. (2011) derive how FEVD can be carried out in one step, making it easier to compute standard errors. Our untangling approach is more clear-cut, as estimates are a linear transformation of standard (zero-normalized one-step) estimates, as Section 4 explains.

A second practicality stems from the fact that FEVD (and the related estimators mentioned earlier) has only been defined for models with α_i effects. We have not encountered generalizations to $\tau_i{\cdot}t$ and θ_t effects. This could be related to the non-trivial specification of the regression in step two. What should be on its left-hand side: within-groups residuals, containing the information on all effects together, or estimated fixed effects such as $\widehat{\theta}_t$, depending on the normalization? Should the right-hand side include a trend besides the intercept, and should the impacts of all constant regressors (v_i and w_t) be estimated jointly? What are the imposed orthogonality restrictions regarding $\mathbb{E}\left\{v_i\tau_i\right\}$? How to obtain proper standard errors? In general, the more extensive the model, the more such questions arise, and the choices matter for the estimates. There may be an acceptable answer, but the point is that it is easy to make mistakes. Untangling avoids such choices and easily handles a broad set of FE configurations.

A third practical difference is that our method yields estimates of the FE α_i^u with standard errors for each i.[2] This helps to test how much and where the model fails to explain in the i-dimension and facilitates the search for new regressors. This advantage remains if $\tau_i \cdot t$ and θ_t effects are added.

4 Estimation of untangled parameters

Consider mean equation (2). Normalizations only affect the lower-dimensional part of the regressors, so the focus in this section is on $D\delta + Z\gamma$. We introduce two methods to estimate the untangled parameters. The indirect approach estimates the model using zero normalization and then renormalizes the estimates into untangling normalization. The direct approach estimates a model with transformed regressors that incorporate the untangling normalization.

4.1 Applicability

Before discussing the approaches in detail, we emphasize their broad applicability in the following senses. They work for any normalization, as long as it is linear in the parameters. We call this a general normalization. The untangled parameters then follow as a special case. Moreover, the methods do not depend on the error term, so that they can be combined with various assumptions and estimators.

It is important, however, that the estimator yields estimates for $D\delta+Z\gamma$. More precisely, this is required for the part of $D\delta+Z\gamma$ that varies over the dimension of interest. For example, if one is interested in the impact of i-regressors, then one must have an estimate of the variation over i, $\alpha_i + v_i'\nu$, so that untangling can use that to derive the estimates of ν^u and α_i^u. Here α, τ, τ_i, θ_t, and ω are nuisance parameters and for them no estimates are required. The least squares dummy variable (LSDV) estimator is a popular option, where the nuisance parameters can be projected out.

The requirement to have estimates of the variation over a specific dimension

[2]For a panel with α_i effects only, the point estimates of the impact of the constant regressor v_i from FEVD and untangling are the same, reflecting that both methods here use the same orthogonality relations (as moment conditions and as normalization, respectively).

calls for ample observations in the other dimension. In our example, the observations of country i over time are used to estimate $\alpha_i + v_i'\nu$, so reliable estimates of it require T to be sufficiently large.

Let us study (small-sample) bias. For notational convenience, do as if $\nu = 0$. First, take the case where the regressors are strictly exogenous regarding the error. Then LSDV is an unbiased estimator of the FE, irrespective of T.

For alternative error assumptions, consider Fernández-Val and Weidner (2017), who allow for predetermined regressors and study linear and the most commonly used nonlinear models. They review the literature on large-N and large-T approximations and conclude that the order of the bias in the asymptotic approximation corresponds with the inverse of the number of observations per parameter. For the estimator of α_i this means $1/T$ (same for τ_i, and $1/N$ for θ_t).

As the authors write, this is useful to get an initial idea of the relevance of bias, but the exact magnitude still depends a lot on the application at hand. For an ARX model with only α_i FE, Buddelmeyer et al. (2008) use simulations to study the bias in the estimated FE. They confirm that the bias drops if T increases. For example, LSDV has a bias of 7% for $T = 10$ and 3% for $T = 20$.[3] This is in line with Fernández-Val and Weidner (2017). Both papers also show how bias correction can further improve small-sample properties. All this suggests that untangling can be fruitfully applied in various practically relevant settings.

4.2 Normalizations in matrix notation

We will set out both estimation approaches in matrix notation, so let's first express the normalizations in that form as well. The general normalization, indicated by g, is represented by a matrix N^g with $m_d + m_z$ independent rows, where each row specifies one normalization, which is a linear combination of δ and γ. That is, the general-normalized parameters δ^g and γ^g obey[4]

$$N^g \begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = 0. \tag{11}$$

[3]Their model has $y_{i,t-1}$ and a strictly exogenous x_{it}. The reported bias is the absolute bias, averaged over i and as a percentage of the standard deviation of α_i, for an autoregressive parameter of 0.4 and $N = 20$. The bias increases with that parameter, but N hardly matters.

[4]One can generalize (11) by allowing the right-hand side to be any non-zero vector c^g. All subsequent expressions can be easily changed accordingly.

This normalization contains zero and untangling normalizations as special cases.

Zero normalization sets elements of δ and γ to zero, denoted with a 0 superscript. We represent it by a matrix N^0, where each row has a one at the place corresponding to the zero-normalized parameter, and for this matrix (11) applies with g substituted by 0. For example, if we normalize the constant to zero, so $\alpha^0 = 0$, then N^0 contains the row $[1, 0, .., 0]$. Instead, if we normalize the i-th country FE to zero, the row is $[0, .., 0, 1, 0, .., 0]$, where the $1 + i$-th element is one.

Untangling normalization is unique. It is based on (5)-(10), represented by N^u. As there are no normalizations on α, τ, and γ, the corresponding columns in N^u are zero.

$$
\begin{array}{c}
\\
N^u =
\end{array}
\begin{array}{c}
\begin{array}{ccccccc} \alpha & \alpha_1 ... \alpha_N & \tau & \tau_1 ... \tau_N & \theta_1 ... \theta_T & \nu' & \omega' \end{array} \\
\begin{bmatrix}
0 & 1...1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1...1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1...1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1...T & 0 & 0 \\
0 & v_1 ... v_N & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & w_1 ... w_T & 0 & 0
\end{bmatrix}
\end{array}
\begin{array}{l}
\text{Row implements:} \\
\sum_i \alpha_i^u = 0 \\
\sum_i \tau_i^u = 0 \\
\sum_t \theta_t^u = 0 \\
\sum_t \theta_t^u \cdot t = 0 \\
\sum_i \alpha_i^u v_i^k = 0 \\
\sum_t \theta_t^u w_t^k = 0.
\end{array}
\tag{12}
$$

4.3 Indirect estimation: renormalizing zero-normalized estimates

The indirect estimation method is based on two steps. It first estimates (2) using zero normalization and then renormalizes the zero-normalized estimates to untangled ones.

4.3.1 Estimating zero-normalized parameters

Using zero normalization in the estimation step is convenient because the parameters can then be estimated in a standard way, as one just omits the variables corresponding to the normalized parameters from the regressor matrix. After estimation, add zeros to the estimated parameter vector and rows and columns of zeros to the estimated covariance matrix corresponding to the zero-normalized parameters. We now have estimates for δ^0 and γ^0 and the corresponding full covariance matrix.

4.3.2 Renormalizing the estimates

General normalization

The second step renormalizes δ^0 and γ^0 into the general-normalized parameters δ^g and γ^g. That requires $K_d + K_z$ independent equations, obtained as follows.

Renormalization is just a redistribution of the total $D\delta^0 + Z\gamma^0$ over $D\delta^g + Z\gamma^g$ where the value for each observation stays the same. This gives NT equalities. But only $K_d + K_z - m_d - m_z$ of them are independent, because that is the column rank of the regressor matrix $[D, Z]$, as shown in Section 2. We get $m_d + m_z$ additional equations by equating (11) with the similar version for the zero normalization. This yields a system containing $K_d + K_z$ independent equations:

$$\begin{bmatrix} D & Z \\ \multicolumn{2}{c}{N^g} \end{bmatrix} \begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = \begin{bmatrix} D & Z \\ \multicolumn{2}{c}{N^0} \end{bmatrix} \begin{bmatrix} \delta^0 \\ \gamma^0 \end{bmatrix}. \tag{13}$$

To solve for δ^g and γ^g, we define

$$R^g = \begin{bmatrix} D & Z \\ \multicolumn{2}{c}{N^g} \end{bmatrix} \text{ and } R^0 = \begin{bmatrix} D & Z \\ \multicolumn{2}{c}{N^0} \end{bmatrix}, \tag{14}$$

and pre-multiply (13) with $R^{g\prime}$. As R^g has full column rank, we obtain

$$\begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = R^{0g} \begin{bmatrix} \delta^0 \\ \gamma^0 \end{bmatrix}, \tag{15}$$

where $R^{0g} = \left(R^{g\prime} R^g\right)^{-1} R^{g\prime} R^0$ is the renormalization matrix that converts the zero into the general normalization, which only consists of observables.[5]

Hence, to obtain the g-normalized estimates we take the zero-normalized estimates and covariance matrix and apply (15). No additional estimation or standard error correction is needed, which is convenient from a practical viewpoint.

Special case: untangling normalization

To estimate the untangled normalized parameters, we use (15) with N^g in R^{0g} replaced by the N^u matrix presented in (12).

---

[5]Computing R^{0g} depends on multiplications involving R^g and R^0, which have many rows. This can be simplified as follows. First, select all $K_d + K_z$ independent rows in R^g by Gaussian elimination, making the resulting $\widetilde{R}^g$ a square matrix of full rank. To maintain the equalities in (13), we then select the same rows in R^0 and obtain the square matrix $\widetilde{R}^0$. Finally, use $\widetilde{R}^{0g} = \widetilde{R}^{g-1} \widetilde{R}^0$ instead of R^{0g} in (15). This yields the same g-normalized parameters.

4.4 Direct estimation: incorporating normalization into regressors

The second estimation method transforms the regressors in the $(D\delta + Z\gamma)$-part of (2) such that they incorporate the normalization and that the regressor matrix becomes full column rank. Now we can directly estimate the transformed model and obtain estimates of the normalized parameters. The direct approach is also useful for estimating models under constraints, which we will need in Section 5.

General normalization

For a set of general-normalized parameters δ^g and γ^g, we split off some resultant parameters by writing them as a function of the free parameters based on the normalization. This can be done as follows.

Because N^g has full row rank $m_d + m_z$, we can take $m_d + m_z$ independent columns of N^g and collect them in N_r^g, which is thus invertible. Let P be the column permutation matrix that forms N_r^g and puts the remaining columns in N_f^g in such a way that we keep the initial order of the parameters in both N_r^g and N_f^g. We split δ^g and γ^g accordingly. That is,

$$N^g P = \begin{bmatrix} N_f^g & N_r^g \end{bmatrix} \text{ and } P' \begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = \begin{bmatrix} \delta_f^g \\ \gamma_f^g \\ \delta_r^g \\ \gamma_r^g \end{bmatrix}. \tag{16}$$

So the choice of P determines what are the free and what are the resultant parameters, but P does not affect the normalization itself.

Using normalization description (11), writing N^g as $N^g P P'$, and using (16) gives

$$\begin{bmatrix} \delta_r^g \\ \gamma_r^g \end{bmatrix} = -N_r^{g-1} N_f^g \begin{bmatrix} \delta_f^g \\ \gamma_f^g \end{bmatrix}. \tag{17}$$

Thus the full parameter vector is a function of the free parameters:

$$\begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = F^g \begin{bmatrix} \delta_f^g \\ \gamma_f^g \end{bmatrix}, \tag{18}$$

where

$$F^g = P \begin{bmatrix} I \\ -N_r^{g-1} N_f^g \end{bmatrix}, \tag{19}$$

and the identity matrix I has size $K_d + K_z - m_d - m_z$.

We partition F^g into four blocks $[F_{11}^g\ F_{12}^g;\ F_{21}^g\ F_{22}^g]$ such that (18) yields

$$D\delta^g + Z\gamma^g = (DF_{11}^g + ZF_{21}^g)\,\delta_f^g + (DF_{12}^g + ZF_{22}^g)\,\gamma_f^g. \tag{20}$$

So we have incorporated the g-normalization into the regressor matrix, which has become $[DF_{11}^g + ZF_{21}^g, DF_{12}^g + ZF_{22}^g]$. Hence, we are in a standard setting, where δ_f^g and γ_f^g can be estimated and (18) then gives the estimate of the full vector.[6]

Special case: untangling normalization

Untangling normalization simplifies the F matrix substantially. We know from the specification of N^u in (12) that there are no normalizations on γ^u, so that $\gamma^u = \gamma_f^u$. Therefore, given (18), $[F_{21}^u, F_{22}^u] = [0, I_{K_z}]$. Moreover, (12) shows that the rightmost K_z columns in N_f^u, which refer to γ^u by construction of P, contain only zeros. So the same holds for $N_r^{u-1} N_f^u$. Hence, considering the complete F^u matrix, (19) implies that its rightmost K_z columns consist of zeros except for a block I_{K_z}. The P matrix in (19) permutes the rows such that I_{K_z} ends up at the rows corresponding to the elements of γ^u, that is, the bottom rows. So above those rows, the rightmost K_z columns in F^u contain only zeros. Hence, F^u is block diagonal:

$$F^u = \begin{bmatrix} F_{11}^u & 0 \\ 0 & I_{K_z} \end{bmatrix}. \tag{21}$$

As a result, (20) simplifies to

$$D\delta^u + Z\gamma^u = D^u \delta_f^u + Z\gamma^u, \tag{22}$$

where $D^u = DF_{11}^u$. Z is not transformed, as there is no normalization on γ^u.

5 Testing constraints that identify γ

We are interested in γ, the true value of the impact of the constant regressors The presence of the fixed effects δ makes that we can only estimate γ^u, leaving γ

[6] To see that the new regressor matrix has independent columns, realize that the column space of $[D, Z]$ is $K_d + K_z - m_d - m_z$-dimensional, so the equality in (20) implies the same for the new matrix. Because this dimension equals the number of columns of the new matrix, its columns are independent.

unidentified. But $\alpha_i^u = 0$ implies $\nu^u = \nu$, and $\theta_t^u = 0$ implies $\omega^u = \omega$, as explained in Section 2.3, so that we have constraints that identify (parts of) $\gamma = [\nu', \omega']'$.

This section introduces two tests. The first examines the constraint directly, so it is a diagnostic test. The second verifies whether the estimates of other parameters are affected by the constraint, so we call this the sensitivity test. Both tests are special cases of the following more general testing procedure.

5.1 Testing under normalizations

As long as untangled parameters are not involved in normalizations, such as α^u and ν^u, we can use a standard way to test constraints, such as a Wald test. But untangled parameters that are normalized, such as α_i^u, are linked to each other and their estimator can have a singular covariance matrix, invalidating standard testing. Things change if we incorporate the normalization into the constraint.

Consider the null hypothesis

$$H_0 : C \begin{bmatrix} \delta^u \\ \gamma^u \end{bmatrix} = c, \tag{23}$$

where C is the matrix of constraints with independent rows, and c is a vector with constraint values.

Testing this involves two problems. First, the estimator of $[\delta^{u\prime}, \gamma^{u\prime}]'$ has a singular covariance matrix, due to the normalization. This is resolved by substituting (18) into (23), giving $CF^u \left[\delta_f^{u\prime}, \gamma_f^{u\prime}\right]' = c$, where the vector of free parameters can be estimated in the standard way with a non-singular covariance matrix (see Section 4.4).

The second problem is that rows in C may be redundant due to the normalization. For example, constraining all α_i^u makes at least one row redundant, because the constraint on α_N^u follows from the constraints on the other α_i^u and the normalization. More formally, CF^u may have dependent rows. We remove those from CF^u and denote the result by C^u. Taking out the corresponding rows from c yields c^u. We thus rewrite

$$H_0 : C^u \begin{bmatrix} \delta_f^u \\ \gamma_f^u \end{bmatrix} = c^u. \tag{24}$$

This can be tested in a standard way, where a Wald test would be a typical choice.

5.2 Diagnostic test

As argued above, the untangled FE α_i^u and θ_t^u provide a way to test for identification. They represent i-variables and t-variables relevant for y_{it} but omitted from the model at hand. Our first test is about existence of such omitted variables, so the null hypothesis constrains α_i^u and/or θ_t^u to zero. This means that a part of δ^u is constrained, and we denote that part by a subscript 0. So the null hypothesis is $\delta_0^u = 0$. This is a special case of (23). Therefore, defining C accordingly, setting $c = 0$, and following the approach of the previous section gives a test. This is our diagnostic test.

The constraint is sufficient for identification of ν and/or ω. It is not necessary, because α_i^u and/or θ_t^u may be uncorrelated with the included regressors, ensuring identification even if they are nonzero. Hence, the diagnostic test may reject even if the true values are identified.

5.3 Sensitivity test

The second test avoids the stringency of the diagnostic test by accounting for the fact that, even if omitted variables exist, they need not matter for estimating parameters of interest. This is similar to the idea underlying an omitted variables bias test. For β as parameter of interest, we thus compare the unconstrained estimator $\widehat{\beta}$ to the estimator $\widetilde{\beta}$ obtained under $\delta_0^u = 0$, which is equivalent to setting its free part $\delta_{f0}^u = 0$.[7] If the estimator of β is insensitive, that is reassuring. If besides this insensitivity in the model with constant regressors, the estimator is sensitive in the model without them, it is the constant regressors that are responsible for the improvement, so they explain an important part of the FE. Note that we do not have such an approach for γ, because we have no unconstrained estimator for it.

We can compute $\widehat{\beta} - \widetilde{\beta}$, but we do not know its variance, as $\widehat{\beta}$ and $\widetilde{\beta}$ are correlated. However, if we focus on LSDV, $\widehat{\beta} - \widetilde{\beta}$ can be written as a transformation

[7] The difference between δ_0^u and δ_{f0}^u is the resultant part δ_{r0}^u. The latter follows from the free part using (17) for $g = u$. Consider the right-hand side of that equation. The normalizations that involve the constrained FE are specific rows in N^u of (12). The rows of the N_f^u-part have zeros at the positions corresponding to all unconstrained FE and γ. The nonzeros in the rows are multiplied by δ_{f0}^u, which is constrained to 0. Hence, the elements in $N_f^u[\delta_f^{u\prime}, \gamma_f^{u\prime}]'$ that correspond to these normalizations are zero. Also the same rows in N_r^u have zeros at the positions corresponding to all unconstrained FE and γ. Hence, (17) implies that δ_{r0}^u is a multiple of only zeros in $N_f^u[\delta_f^{u\prime}, \gamma_f^{u\prime}]'$, so that $\delta_{r0}^u = 0$.

of $\widehat{\delta}^u_{f0}$, based on Magnus and Vasnev (2007):

$$\begin{bmatrix} \widehat{\delta}^u_{f\emptyset} \\ \widehat{\gamma}^u \\ \widehat{\beta} \end{bmatrix} - \begin{bmatrix} \widetilde{\delta}^u_{f\emptyset} \\ \widetilde{\gamma}^u \\ \widetilde{\beta} \end{bmatrix} = -\left([D^u_\emptyset, Z, X]' \, [D^u_\emptyset, Z, X]\right)^{-1} [D^u_\emptyset, Z, X]' \, D^u_0 \widehat{\delta}^u_{f0}, \qquad (25)$$

where $\delta^u_{f\emptyset}$ collects the elements of δ^u_f that are not in δ^u_{f0}, and $D^u_\emptyset$ is the corresponding submatrix of D^u used in (22). We know the distribution of $\widehat{\delta}^u_{f0}$ and thereby of $\widehat{\beta} - \widetilde{\beta}$. So we can test whether its realization differs significantly from zero. This is our sensitivity test. It essentially takes (24) and uses a specific linear combination of δ^u_{f0}, illustrating that not the mere absence of omitted variables ($\delta^u_{f0} = 0$) is crucial, but rather how much a combination of them matters for the estimating parameters of interest.

To interpret (25), distinguish two parts on the right. At the end, we have the diagnostic part, $\widehat{\delta}^u_{f0}$, which measures the magnitude of the misspecification due to constraining FE to zero. The remainder indicates how much one unit of misspecification matters for the estimate of β (and the other parameters), so it is a derivative. Even a large and/or significant $\widehat{\delta}^u_{f0}$ can barely matter for estimating β, if the derivative in that direction is low. Magnus and Vasnev (2007) emphasize the importance of analyzing the derivative in addition to the diagnostic. Our sensitivity test accounts for both aspects.

6 Application: untangling the gravity model

One can apply the untangling normalizaton method to various economic models to try to identify parameters that were thought to be unidentifiable due to the added FE. Take for example the gravity model.

6.1 The gravity model and the identification problem

The general idea of the gravity model is that exports between two countries depend positively on both the exporting and the importing country's GDP and negatively on world GDP and the distance between the countries, where distance can be both physical and economic distance, such as trade costs.

The most common gravity model is that of Anderson and van Wincoop (2003). They show that, besides bilateral trade costs, it is important to include multilateral resistance terms for the importer and exporter to avoid estimation bias. It has been difficult, however, to find economic variables to capture these terms.

As gravity models are typically analyzed using data on country pairs over time, one could nevertheless control for the multilateral boundaries by country-time FE. But then the underlying dummies are perfectly multicollinear with country- and time-specific variables, such as exporter, importer, and world GDP, so that the impacts of a variety of potentially interesting export determinants are not identified. This is a well-known problem, and authors have tried to circumvent it, as Head and Mayer (2014) describe. These approaches are examples of imposing the RE-type of restrictions discussed in Section 3.4.

As an alternative, we work in a fully FE-setting and apply our untangling normalization method. After estimation, we can test whether the remaining, untangled FE still matter. If not, they can be left out and we have estimated the previously unidentified parameters. That would also mean identification of the relevance of economic variables for the multilateral resistance terms.

6.2 Model specification

Consider exports from country i to the US in year t. Taking one importer is for simplicity and to stay in line with the it setting of previous sections (see Klaassen and Teulings (2017) for a three-dimensional application). The model specifies

$$\begin{aligned} exp_{iUSt} =& \beta_1 gdp_{it} + \beta_2 reer_{it} + \omega_1 gdp_{USt} + \omega_2 gdp_{Wt} + \omega_3 reer_{USt} \\ &+ \alpha + \alpha_i + \tau\cdot t + \tau_i \cdot t + \theta_t + \varepsilon_{it}, \end{aligned} \tag{26}$$

where exp_{iUSt} represents real exports from country i to the US, and gdp_{it}, gdp_{USt} and gdp_{Wt} are real GDP of country i, the US, and the world, respectively, all in constant dollars. Moreover, $reer_{it}$ and $reer_{USt}$ are the real effective exchange rates (REER) of country i and the US, respectively, where exchange rates are defined as home currency units per unit of foreign currency, so an increase in REER means a depreciation. All variables are in log. We thus have two it-regressors, gdp_{it} and $reer_{it}$, and three t-regressors, gdp_{USt}, gdp_{Wt}, and $reer_{USt}$, where the latter are the constant regressors.[8] We also include lags of all regressors but ignore them

[8]The gdp_{Wt} regressor resembles a Mundlak (1978) term, that is, an average of gdp_{it} over i. Mundlak, however, uses averages as auxiliary regressors to control for correlation between his random effect and his it-regressors; he has a correlated random effects approach, so there are no constant regressors. In contrast, world GDP in our model is motivated by economic theory — it is not an auxiliary regressor — and its impact is a parameter of interest. Moreover, we treat the remaining time effect as fixed instead of random, thereby recognizing its correlation

in (26) for simplicity of notation.

The GDP variables are suggested by the typical gravity model. The theory in Klaassen and Teulings (2017) proposes adding real exchange rates, both the bilateral rate rer_{iUSt} and the partners' REERs, that is, $reer_{it}$ and $reer_{USt}$. Because triangular arbitrage implies $rer_{iUSt} = reer_{it} - reer_{USt}$, only the REERs are included as regressors, realizing that β_2 is the impact of $reer_{it}$ plus that of rer_{iUSt}, and ω_3 is the impact of $reer_{USt}$ minus that of rer_{iUSt}.

We add the general set of FE. The exporter effects α_i control for distance between the exporting country and the US, common language with the US, and so on. Exporter-trend effects τ_i account for the strong trends in exports that are not explained by regressors, reflecting the typical approach in time series modeling to avoid that, for example, the GDP parameters are abused to also correct for omitted trending variables; Bun and Klaassen (2007) and Baier et al. (2014) confirm the importance of $\tau_i \cdot t$ in panel gravity estimation. Time effects θ_t capture global and US-specific developments, such as the recent financial crisis, and US dollar swings.

The multilateral import resistance of the US is fully controlled for by $reer_{USt}$, α, $\tau \cdot t$, and θ_t. Regarding the export side, the model cannot perfectly account for the multilateral resistance of country i, as we have no it-FE. However, the included FE correct for all its time-constant, trend, and global variation, and we have $reer_{it}$ as regressor, one of the prominent determinants of the multilateral export resistance, as Klaassen and Teulings (2017) derive theoretically.

We assume a zero mean for the error term ε_{it} conditional on the regressors in all times. We thus ignore feedback from bilateral exports to GDPs and REERs, which is in line with the gravity literature and seems reasonable given that bilateral exports are a limited fraction of total exports and thus GDP and that exchange rates are mainly driven by financial variables. The Wooldridge (2010, p. 325) test support this, as leads of regressors have insignificant impacts. The error term is allowed to be heteroscedastic and serially correlated.

This model, estimated by LSDV, is sufficient to illustrate our method. Our main results are robust against different specifications, such as omitting country-trend FE, accounting for non-stationarity and cointegration, and a multiplicative approach estimated by Gamma and Poisson pseudo maximum likelihood (GPML and PPML), following Santos-Silva and Tenreyro (2006), as Appendix A shows.

---

with the *it*- and constant regressors.

6.3 Data

We focus on $N = 17$ countries, namely the EU-15 countries except for Belgium and Luxembourg, Canada, Japan, Norway and Switzerland. The sample is from 1979-2011 ($T = 33$), resulting in 561 observations.

We use monthly nominal export data from the IMF Direction of trade statistics (DOTS) and convert them back into home currency using the monthly dollar exchange rate from the International Financial Statistics (IFS) of the IMF. We then sum to get yearly values and divide by the home export price index from the European Commission AMECO database (the base year for all data is 2010). We divide by the home PPP of the dollar from the OECD Economic Outlook to obtain exports in constant dollars.

Nominal yearly GDP data are from AMECO, and we use the AMECO exchange rate to express it in national currency. We then divide by the AMECO GDP deflator and by the home PPP of the US dollar to get GDP in constant US dollars. West-German data is used as a proxy for Germany before 1991. We take real world GDP in US dollars from the OECD Economic Outlook.

Finally, we use consumer-price-based monthly REER data from the Bank for International Settlements (BIS), construct yearly averages, and invert.

6.4 Zero normalization and estimation

We estimate model (26) using the indirect approach of Section 4.3. That is, we first impose a zero normalization and estimate the model, using LSDV. In the next section we renormalize into the untangling normalization.

Table 1 specification 4 presents the estimation results. In the zero normalization, $\omega^0 = 0$, so we only have the two it-regressors. Their estimates do not depend on the normalization. The signs can be explained by the theoretical gravity model. If exporter's GDP increases, the exporting country produces more goods, reducing the price and thereby increasing exports to the US.

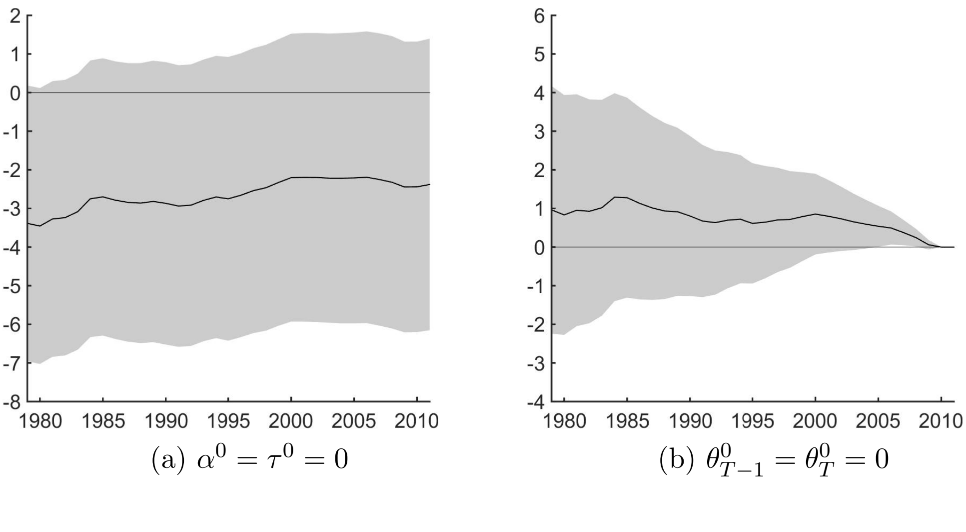
The parameter for $reer_{it}$ captures the impacts of the bilateral rate rer_{iUSt} and the multilateral $reer_{it}$, as explained in Section 6.2. Consider an increase in $reer_{it}$. First, this depreciates the currency of i against the US, making country i cheaper for the US, so it will export more to the US. Second, the increase in $reer_{it}$ means that the currency of i depreciates against all currencies in the rest of the world (RoW) as a whole. Therefore, country i becomes cheaper for RoW and the demand for its products increases, raising prices, so that its exports to the US

Table 1: Estimation results for exp_{iUSt} based on model (26)

Include θ_t	No			Yes		
Specification	1	2	3	4	5	6
Description	no w_t	w_t	w_t,DL	θ_t, no w_t	θ_t,w_t	θ_t,w_t,DL
gdp_{it}	2.09 *	0.62 *	1.13 *	0.80 *	0.80 *	1.12 *
	(0.29)	(0.23)	(0.24)	(0.24)	(0.24)	(0.25)
$reer_{it}$	1.24 *	0.38 *	0.68 *	0.42 *	0.42 *	0.66 *
	(0.18)	(0.12)	(0.14)	(0.12)	(0.12)	(0.15)
gdp_{USt}		3.17 *	2.30 *		3.03 *	2.35 *
		(0.29)	(0.37)		(0.29)	(0.39)
gdp_{Wt}		−1.79 *	−1.23		−1.84 *	−1.18
		(0.48)	(0.72)		(0.45)	(0.73)
$reer_{USt}$		−1.04 *	−1.01 *		−1.02 *	−1.02 *
		(0.11)	(0.13)		(0.10)	(0.13)
Wald tests						
$\theta_t^u = 0$	–	–	–	16.11 *	3.05 *	1.14
				[0.00]	[0.00]	[0.31]
$\beta = \beta\|_{\theta_t^u=0}$	–	–	–	138.54 *	4.96 *	0.33
				[0.00]	[0.01]	[0.72]
R_θ^2	–	–	–	0	0.91	0.98

Static models have 561 observations and distributed lag (DL) models 527. Both DL models have two lags of each regressor, and we display the long-run effects.
The first Wald statistic has $T - m_\theta - K_w$ constraints, where m_θ is the number of normalizations to untangle θ_t from the other FE. For Model 5 this gives 28. The second Wald, denoted by $\beta = \beta|_{\theta_t^u=0}$, is the test of Section 5.3, showing how sensitive the estimator for β of the level variables is to setting $\theta_t^u = 0$ and therefore has 2 constraints. Both Wald tests are divided by the number of constraints, and we relate them to an F-distribution. It has $NT - K_x - (K_d - m_d)$ denominator degrees of freedom, which is 494 for Model 5.
R_θ^2 is the fraction of the variance of the untangled time FE from a model without t-regressors that is explained once (detrended) t-regressors are included.
Standard errors are between brackets and they are based on Newey and West (1987, 1994), which gives three lags. p-values are in square brackets. * indicates significance at the 5% level, the level that we use throughout the paper.

Figure 1: Zero-normalized time FE θ_t^0 and their dependence on the normalization.



(a) $\alpha^0 = \tau^0 = 0$ (b) $\theta_{T-1}^0 = \theta_T^0 = 0$

fall. The estimate shows that the bilateral effect dominates the multilateral one.

The estimates for the time FE θ_t^0 depend on the chosen zero normalization. We study two of them. Both have $\alpha_N^0 = \tau_N^0 = 0$ and $\omega^0 = 0$, but they differ regarding the other two parameters.

Figure 1a shows the estimated time FE and confidence band for $\alpha^0 = \tau^0 = 0$. Their mean is negative, reflecting that they are disturbed by the overall means of the dependent and explanatory variables. They exhibit some variation over time, but this seems a minor feature. They are jointly significant, as the Wald test of $\theta_t^0 = 0$ for all t (divided by the number of constraints) is 18.81, with p-value 0.00.

Figure 1b displays the results for the normalization $\theta_{T-1}^0 = \theta_T^0 = 0$. This leads to somewhat downward sloping estimates with a positive mean. This change compared to Figure 1a exemplifies the impact that different zero normalizations have on the FE estimates, which hampers their interpretation. The Wald test of 16.11 (p-value 0.00) is also different.

6.5 Untangling normalization and identifying ω

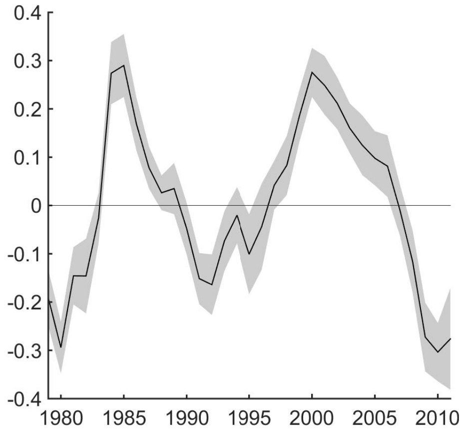
The zero-normalized time FE in Figure 1 are difficult to interpret. In Section 6.5.1 we show how untangling normalization improves on that. Section 6.5.2 introduces t-regressors and shows they explain a large part of the time FE. Finally, in Section 6.5.3 we add distributed lags (DL) to the model, leading to our baseline specification and the answer to the question whether we can identify the impacts of the t-regressors.

6.5.1 Static model without t-regressors

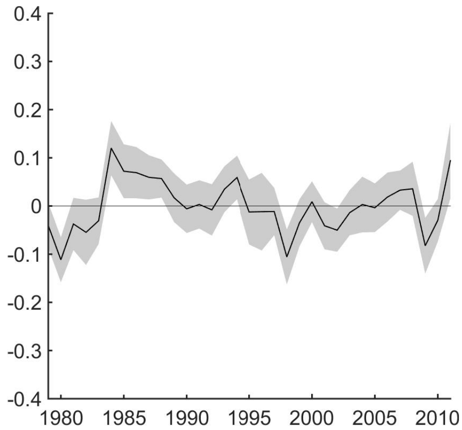
For the moment, we keep out the t-regressors ($\omega = 0$) and just redistribute the FE estimates obtained from some zero normalization into the uniquely determined untangled FE estimates, using (15) and (12). The untangled time FE θ_t^u are thus cleaned from the overall level and trend, and for the rest they are freely determined by the data.

Figure 2a illustrates the first main contribution of untangling normalization: the estimated FE are more informative and easier to interpret compared to the zero-normalized ones in Figure 1. We now recognize a business cycle. Notice the dot-com bubble and, clearly visible, the recent financial crisis. We also see dollar exchange rate effects, such as the dollar bubble in the eighties, stimulating and then hampering exports to the US.

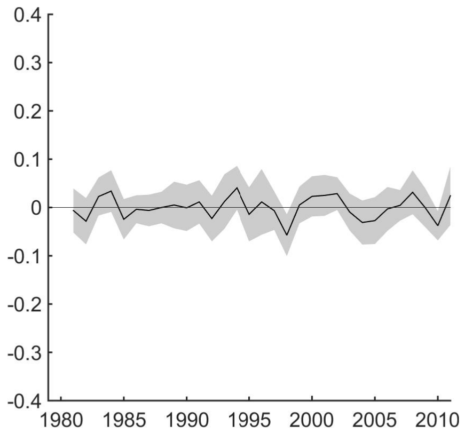
Figure 2: Untangled time FE θ_t^u for models without and with t-regressors.



(a) Static Model 4, without t-regressors.



(b) Static Model 5, with t-regressors.



(c) DL Model 6, with t-regressors.

6.5.2 Static model with t-regressors

The theoretical gravity model supports the insights just drawn from Figure 2a, as it argues that the t-regressors US GDP, world GDP and US REER matter for exports to the US. So we now add these to the model by leaving ω free.

Figure 2b shows a striking difference with Figure 2a. The estimated θ_t^u that remain when using the three t-regressors are much closer to zero. So the regressors explain most of the time-specific variation, quantified by $R_\theta^2 = 91\%$, which is defined in the note to Table 1. Therefore, exploiting constant regressors, which untangling can but traditional (zero) normalization cannot do, is a promising step to achieve identification of the true value of ω. But there is room for further improvement, as the Wald diagnostic and sensitivity tests reject that the θ_t^u can be left out; see Table 1 Model 5.

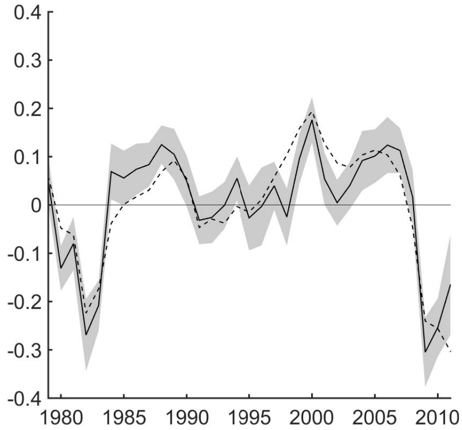
So far, we have looked at the explanation of the t-regressors jointly. To examine the relevance of US GDP individually, we re-estimate Model 5 while leaving US GDP out. The resulting $\hat{\theta}_t^u$ are represented by the solid black line in Figure 3a. The dashed line is US GDP (transformed to facilitate comparison; see the figure note for details). The resemblance is striking. The estimated θ_t^u almost exactly mimic the US business cycle. Hence, US GDP is important in explaining time FE, which is supported by the high partial $R_\theta^2 = 81\%$. For world GDP in Figure 3b the resemblance is rather weak, with a partial $R_\theta^2 = 27\%$, but US REER in Figure 3c explains a lot, with a partial $R_\theta^2 = 74\%$.

6.5.3 Dynamic model with t-regressors

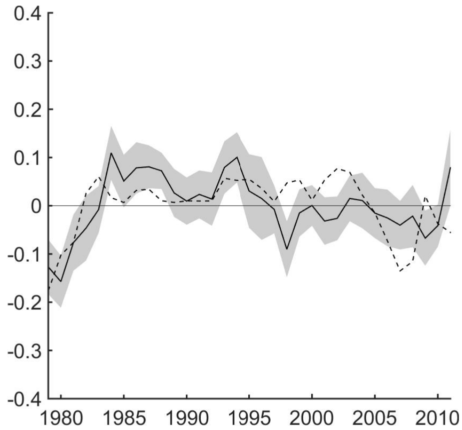
Until now, we have used static models. But traders often entered into contracts in previous periods to export goods in period t, based on export determinants back then. We thus add lagged regressors, that is, lags of it-regressors to take some noise out of θ_t^u and lagged t-regressors to further explain θ_t^u. Lags of unobserved export determinants are still left in the FE (and error term), so all unobserved time-specific developments are grouped in θ_t^u and can help their interpretation, in agreement with the idea of untangling. We thus take a distributed lag (DL) model, where we allow for unrestricted serial correlation in the error.

Two lags of the five regressors turn out to be sufficient, and we add the lags in the form of first differences, for example Δgdp_{it} and $\Delta gdp_{i,t-1}$. This defines Model 6. We focus on the long-run effects, that is, the parameters of the level regressors; the results for the first differences do not alter our conclusions.

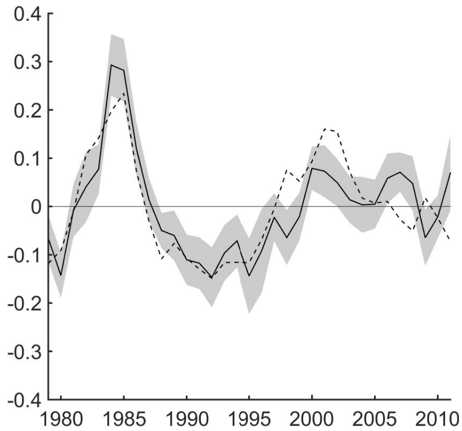
Figure 3: Untangled time FE θ_t^u and its explanation by one t-regressor in Model 5.



(a) Model with gdp_{Wt} and $reer_{USt}$. Dashed line: gdp_{USt}.



(b) Model with gdp_{USt} and $reer_{USt}$. Dashed line: gdp_{Wt}.



(c) Model with gdp_{USt} and gdp_{Wt}. Dashed line: $reer_{USt}$.

The subcaption shows which two t-regressors are included to estimate θ_t^u (solid line). The other t-variable is plotted (dashed line) after detrending, demeaning, and scaling to facilitate comparison. That is, the detrended and demeaned t-variable is multiplied by the ratio of the standard deviations of $\hat{\theta}_t^u$ over that of the transformed t-variable. World GDP and US REER are also multiplied with minus one (Figure 3b and 3c, respectively).

Figure 2c shows the estimated θ_t^u for the DL model. They are close to zero. Hence, we can explain almost all of the $T = 33$ time FE by just three t-regressors and their lags, which is quantified by $R_\theta^2 = 98\%$.

Formal tests in Table 1 Model 6 confirm that we can safely leave out the time FE. After all, the Wald diagnostic test does not reject that all $\theta_t^u = 0$.[9] The Wald sensitivity test shows that leaving out the time FE (that is, using Model 3) does not significantly alter the two β estimates, so there is no evidence of omitted variable bias.

We conclude that untangling normalization and three t-regressors have made the remaining time FE redundant. Our estimates for the impacts of the t-regressors thus reflect their true values instead of only pseudo-true values. This is remarkable, because identifying such true values has been a notorious problem in the literature.

Of particular interest for the gravity literature is that the redundancy of θ_t^u identifies the multilateral import resistance for the US up to a constant and trend, as explained in Section 6.2, with the $reer_{USt}$ as key determinant. This also suggests that $reer_{it}$ is an important determinant of the multilateral export resistance of country i.

The signs of the three ω^u estimates can be explained by the theoretical gravity model (as holds for the β estimates, discussed in Section 6.4). US GDP has a positive effect on exports to the US, because higher income for the US means higher demand for goods and thus imports.

An increase in world GDP has a negative impact because of two reasons. First, the demand for goods of country i by RoW increases, leading to a higher price for its goods, which reduces its exports to the US. Second, RoW produce more goods making them cheaper and therefore the US substitutes its imports from country i to RoW.

The final variable, US REER, reflects the influence of the bilateral rate rer_{iUSt} and the multilateral $reer_{USt}$, as explained in Section 6.2. Suppose $reer_{USt}$ increases. First, this appreciates the currency of i against the US, resulting in

[9] As predicted by Section 2.3, the Wald value is not specific to our choice of using untangling normalization for θ_t, as long as we only normalize parameters that also appear in the null hypothesis. Indeed, if we normalize $\theta_{T-10}^0 = \ldots = \theta_T^0 = 0$, we get exactly the same Wald, while the normalizations of Figures 1a and 1b, $\alpha^0 = \tau^0 = 0$ with $\omega^0 = 0$ and $\theta_{T-1}^0 = \theta_T^0 = 0$ with $\omega^0 = 0$, respectively, give Wald values of 17.34 and 14.47 (p-values 0.00). Hence, avoiding the redundant additional constraints that the null implies for the latter two normalizations (for example, that GDP plays no role for exports) can be important in practice.

lower exports to the US. Second, the increase in $reer_{USt}$ means that the dollar depreciates against all currencies in the RoW, making it more expensive for the US to import from the RoW and as a consequence the US imports more products from country i. The negative estimate for $reer_{USt}$ shows that the bilateral effect dominates the multilateral one, as it did for $reer_{it}$.

6.6 Enlarged sample: 1965-2011

So far, we have considered a sample from 1979-2011. That has been sufficient to illustrate how untangling normalization works, and how it helps interpretation and identification. Now, and only in this section, we add pre-1979 data, thereby including some economically unstable years.

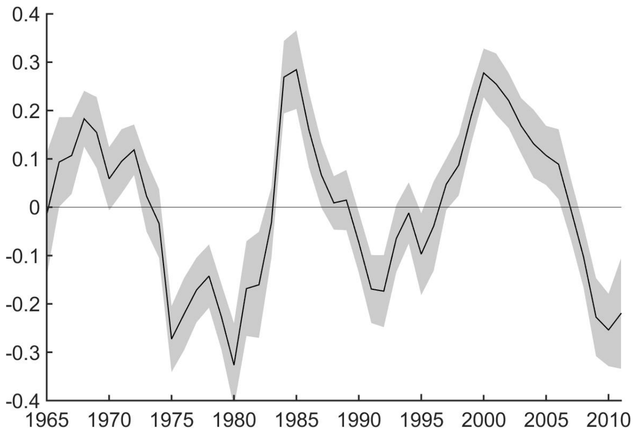
Table 2: Estimation results for exp_{iUSt} model (26) using the enlarged sample

Include θ_t	No	Yes
Specification	3E	6E
Description	DL	θ_t,DL
gdp_{it}	1.43 *	1.39 *
	(0.18)	(0.20)
$reer_{it}$	0.30 *	0.28 *
	(0.12)	(0.12)
gdp_{USt}	2.15 *	2.27 *
	(0.38)	(0.40)
gdp_{Wt}	−1.72 *	−1.67 *
	(0.47)	(0.47)
$reer_{USt}$	−1.26 *	−1.26 *
	(0.11)	(0.10)
Wald tests		
$\theta_t^u = 0$	–	2.43 *
		[0.00]
$\beta = \beta\|_{\theta_t^u=0}$	–	0.39
		[0.68]
R_θ^2	–	0.95

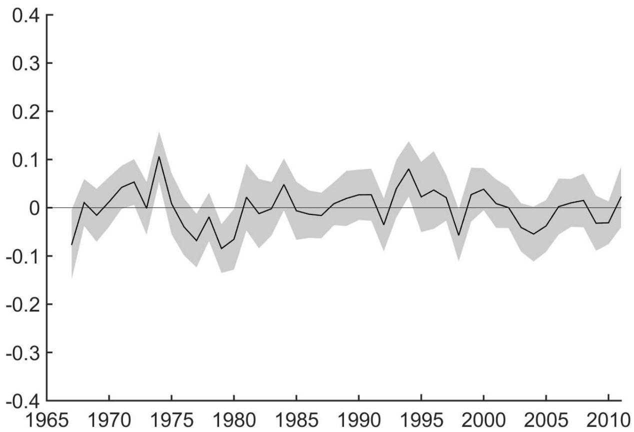
Both models have two lags for every regressor and use data of 1965-2011, resulting in 765 observations. The note to Table 1 provides further details.

Table 2 specification 6E displays the estimation results for the DL model. The main difference with before is that the Wald test now rejects that all untangled time FE $\theta_t^u = 0$, so that we cannot conclude that the estimated ω^u reflects the true value. However, Figure 4b shows that the estimated θ_t^u are still close to zero, reflecting that the time FE are almost completely explained by the t-regressors ($R_\theta^2 = 95\%$). Even the big economic swings before 1979, made visible by untangling in Figure 4a, are captured quite well by the t-regressors. The significance of the Wald test is mainly due to a few significant FE in the 1970s. Moreover, nonzero time effects might be uncorrelated with the included regressors such that ω^u can still equal the true value. This illustrates the stringency of our Wald diagnostic test; it is sufficient but not necessary for $\omega^u = \omega$.

Figure 4: Untangled time FE θ_t^u for the enlarged sample: without and with t-regressors.



(a) Static model, without t-regressors.



(b) DL Model 6E, with t-regressors.

In contrast to the diagnostic test, the Wald sensitivity test does not reject. That is, there is no evidence that the t-variables driving θ_t^u are correlated with the two it-regressors. The latter are similar to the included t-regressors, as both concern GDP and REER. This suggests that leaving out θ_t^u does not cause omitted variable bias in the estimated ω^u as well. This corroborates our qualifications regarding the diagnostic test rejection above. Furthermore, all estimates are quite similar to those in the baseline sample, where we do not reject $\theta_t^u = 0$. So, despite the rejection of the diagnostic test, we tentatively conclude that also in the enlarged sample the estimated ω^u reflect the true value acceptably well.

7 Conclusion

We have shown that in fixed-effects models the true values of the impacts of constant regressors are identified if normalized FE are zero. The reasonableness of this identifying constraint depends on the normalization used to avoid multicollinearity. We have introduced a new normalization method: untangling normalization. It disentangles FE-types from each other and from constant regressors. This not only eases the FE interpretation substantially, but also makes the identifying constraint more realistic, thereby facilitating the identification of the impacts of constant regressors, a notorious problem in the literature.

The untangled parameter estimates and covariance matrix are just linear transformations of their zero-normalized counterparts and can be obtained without an extra estimation step. Alternatively, regressors can be transformed to incorporate the normalizations so that the parameters can be estimated directly.

We have applied untangling normalization to a panel gravity model for exports to the US. This illustrates the two advantages of untangling normalization, regarding interpretation and identification. First, the estimated untangled time FE reveal a business cycle pattern together with a dollar exchange rate effect. Together with gravity theory, this yields three t-regressors — US GDP, world GDP and US REER — to explain the time-specific developments in exports.

Second, we obtain estimates of the true values of the impacts of those three t-regressors, even though that is typically considered beyond reach. Even more so, we identify the US effective exchange rate as the driver of detrended US multilateral import resistance, a variable that has been considered unobservable in the gravity literature so far.

This paper is about two-dimensional panels, with dimension it. One idea for future work is to extend untangling normalization to higher-dimensional panels. For example, a three-dimensional ijt generalization would facilitate the study of financial or trade relations involving many sectors j. Moreover, untangling normalization illustrates the value of the information in estimated FE, which may stimulate further research on their estimation.

A Appendix: robustness analysis

This appendix confirms that our results are robust to various deviations from the baseline specification, such as leaving out country-specific trends, accounting for the non-stationarity of some variables, and a multiplicative model.

Table A.1: Sensitivity of results for exp_{iUSt} model (26)

Specification	6	7	8	9	10
Estimation	LSDV	No $\tau_i \cdot t$	DOLS	GPML	PPML
gdp_{it}	1.12 *	2.26 *	0.46	1.12 *	0.25
	(0.25)	(0.19)	(0.33)	(0.25)	(0.36)
$reer_{it}$	0.66 *	0.83 *	0.50 *	0.67 *	0.36 *
	(0.15)	(0.21)	(0.16)	(0.15)	(0.10)
gdp_{USt}	2.35 *	1.55 *	2.81 *	2.26 *	3.05 *
	(0.39)	(0.50)	(0.33)	(0.39)	(0.38)
gdp_{Wt}	−1.18	−1.06	−1.21	−1.19	−2.46 *
	(0.73)	(1.13)	(0.75)	(0.72)	(0.85)
$reer_{USt}$	−1.02 *	−0.97 *	−1.07 *	−1.01 *	−0.43 *
	(0.13)	(0.17)	(0.13)	(0.13)	(0.18)
Wald tests					
$\theta_t^u = 0$	1.14	1.02	1.15	1.12	2.79 *
	[0.31]	[0.44]	[0.30]	[0.33]	[0.00]
$\beta = \beta\|_{\theta_t^u=0}$	0.33	1.62	0.47	0.27	1.48
	[0.72]	[0.20]	[0.62]	[0.76]	[0.23]
R^2_θ	0.98	0.97	0.98	0.98	0.97

All models have two lags for every regressor. Model 7 leaves out the country-specific trends, that is, $\tau_i = 0$. Model 8 explicitly accounts for cointegration between exp_{iUSt} and gdp_{it} and it uses DOLS estimation with two leads and lags for gdp_{it}. For Models 9 and 10, the Wald sensitivity tests no longer use the transformation in (25), but the corresponding one for the maximum likelihood (ML) estimator, as derived by Magnus and Vasnev (2007). This can be applied to GPML and PPML, because the transformation relies on the first-order condition of ML, which is identical for ML and PML for both the Gamma and Poisson approaches. The note to Table 1 provides further details.

A.1 Leaving out country-specific trends $\tau_i \cdot t$

The number of papers that include trend FE $\tau_i \cdot t$ into a gravity type of model is growing. They are also relevant here, but they do not drive our main results.

More specifically, the Wald test of $\tau_i = 0$ for all i (leaving τ unrestricted) is 276, much higher than the critical value of 26 based on the χ^2_{16}-distribution. Moreover, leaving out country trends affects the estimates, as follows from comparing Specification 7 in Table A.1 to the baseline, replicated as 6. For example, the estimate for gdp_{it} changes, which can be explained by the fact that gdp_{it} is the only i-dependent regressor with a clear trend, so that it will try to fit the omitted country trends as well.

Excluding $\tau_i \cdot t$, however, does not change our main results. The t-regressors still explain most of the time effects ($R^2_\theta = 97\%$), and the two Wald tests confirm that the θ^u_t are jointly insignificant and removing them does not notably affect the β estimates.

A.2 Non-stationarity and cointegration

We first test whether exp_{iUSt}, gdp_{it}, and $reer_{it}$ have a unit root for all countries, using the four Fisher type tests in Stata. The test equation accounts for a drift term, lagged differences, and for exp_{iUSt} and gdp_{it} it also has a trend. All parameters are country specific, and we account for time effects. The results indicate that exp_{iUSt} and gdp_{it} have a unit root, but $reer_{it}$ is stationary.

Next, we apply the Pedroni panel cointegration tests. The test equation contains country effects and trends, and the cointegrating parameter is country specific. We add time effects. There is strong evidence for cointegration between exp_{iUSt} and gdp_{it}.

Although the LSDV estimates used earlier remain consistent in the presence of cointegration, the standard errors require adjustment. We thus perform dynamic OLS (DOLS), as proposed by Mark and Sul (2003). That is, we estimate the cointegrating regression of exp_{iUSt} on gdp_{it}, adding two leads and lags of the first differences of the regressor (combinations of 0-3 leads and lags yield similar results), allowing their coefficients to be country specific, and including the full set of fixed effects. This gives the DOLS estimate of β_1 and its standard error. Then we fix β_1 at this value and estimate the remaining parameters using LSDV.

Model 8 in Table A.1 displays the results. They do not differ much from the baseline results in Model 6. Both Wald statistics do not reject, and the estimated θ^u_t (not shown) are comparable to those in Figure 2c. The time FE are for 98% explained by the t-regressors, in line with the baseline.

A.3 Multiplicative model, estimated by Gamma and Poisson PML

Instead of our constant conditional mean restriction on ε_{it}, motivating LSDV estimation, one may prefer a multiplicative approach by assuming that restriction for exp (ε_{it}) and then use GPML or PPML. Models 9 and 10 of Table A.1 display the results.

The GPML results are close to those of LSDV, so the difference in moment restrictions does not matter for our data. For PPML we reject $\theta_t^u = 0$. However, the plot of estimated θ_t^u is similar to that of GPML (both not shown) and LSDV in Figure 2c, and only five are significant. So we are still close to the LSDV conclusion, which is confirmed by the again high $R_\theta^2 = 97\%$. The Wald sensitivity test does not reject, in line with LSDV.

References

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

Baier, S. L., J. H. Bergstrand, and M. Feng (2014): "Economic integration agreements and the margins of international trade," *Journal of International Economics*, 93, pp. 339–350.

Breusch, T., M. B. Ward, H. T. M. Nguyen, and T. Kompas (2011): "On the Fixed-Effects Vector Decomposition," *Political Analysis*, 19, pp. 123–134.

Buddelmeyer, H., P. H. Jensen, U. Oguzoglu, and E. Webster (2008): "Fixed Effects Bias in Panel Data Estimators," IZA Discussion Papers 3487, Institute for the Study of Labor (IZA).

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

Fernández-Val, I. and M. Weidner (2017): "Fixed Effect Estimation of Large T Panel Data Models," Papers 1709.08980, arXiv.org.

Greene, W. (2011): "Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time-Invariant Variables in Fixed Effects Models?" *Political Analysis*, 19, pp. 135–146.

Hausman, J. A. and W. E. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, pp. 1377–1398.

Head, K. and T. Mayer (2014): *Gravity Equations: Workhorse,Toolkit, and Cookbook*, Elsevier, vol. 4 of *Handbook of International Economics*, chap. 3, pp. 131–195.

Honoré, B. E. and M. Kesina (2017): "Estimation of Some Nonlinear Panel Data Models With Both Time-Varying and Time-Invariant Explanatory Variables," *Journal of Business & Economic Statistics*, 35, pp. 543–558.

Klaassen, F. and R. Teulings (2017): "Untangling Real Gravity," Tinbergen Institute Discussion Papers 17-121/VI, Tinbergen Institute.

Magnus, J. R. and A. L. Vasnev (2007): "Local sensitivity and diagnostic tests," *Econometrics Journal*, 10, pp. 166–192.

Mark, N. C. and D. Sul (2003): "Cointegration Vector Estimation by Panel DOLS and Long-run Money Demand," *Oxford Bulletin of Economics and Statistics*, 65, pp. 655–680.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, pp. 69–85.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, pp. 703–708.

——— (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, pp. 631–653.

Pesaran, M. H. and Q. Zhou (2016): "Estimation of time-invariant effects in static panel data models," *Econometric Reviews*, 35, pp. 1–35.

Plümper, T. and V. E. Troeger (2007): "Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects," *Political Analysis*, 15, pp. 124–139.

——— (2011): "Fixed-Effects Vector Decomposition: Properties, Reliability, and Instruments," *Political Analysis*, 19, pp. 147–164.

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

Suits, D. B. (1984): "Dummy Variables: Mechanics v. Interpretation," *The Review of Economics and Statistics*, 66, pp. 177–180.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, Cambridge MA: The MIT Press.