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

November 15, 2020

Abstract

Fixed effects (FE) in panel data models require normalization 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. We show, however, there is identification if the normalized FE are zero, whatever the normalization. This gives a testable constraint for identification. If it holds, FE can be left out. If not, we propose "untangling normalization" to ease interpretation of the FE and find omitted regressors. In a gravity model for exports to the US, the constant regressors US GDP, world GDP, and US effective exchange rate explain 98% of the time FE, making these FE redundant. We thus achieve identification.

Keywords: gravity model, identification, multicollinearity, normalization, unobserved heterogeneity.
JEL classification: C18; C23; F14.

---

*Corresponding author. Amsterdam School of Economics, PO Box 15867, 1001 NJ Amsterdam, The Netherlands; tel. +31-20-5254191; e-mail: f.klaassen@uva.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, Jan Kiviet, Frank Kleibergen, Maurice Koster, Naomi Leefmans, Jan Magnus, Andrey Vasnev, and seminar/conference participants at the University of Amsterdam, Panel Data Workshop in Amsterdam, 6th Economics & Finance Conference at the OECD Headquarters in Paris, University of Göttingen, 17th Annual Conference of the European Trade Study Group in Helsinki, CPB Netherlands Bureau for Economic Policy Analysis, Tinbergen Institute Amsterdam, and the 25th International Panel Data Conference in Vilnius for very useful and constructive comments.

1 Introduction

In panel data models, researchers often include fixed effects (FE) to control for unobserved heterogeneity. However, many important research questions concern regressors that only vary in the same dimension as the FE. Think of the impact of gender and ethnicity on wage inequality in panels with individual FE, and the effect of geographical characteristics or time-invariant institutions on the economic development of countries in case of country FE. Similarly, one would like to know the impact of global economic and environmental developments on countries, while controlling for time FE.

This paper studies the perfect multicollinearity due to FE and such "constant" regressors (constant, as they are invariant in a dimension of the panel). Multicollinearity requires normalization of some parameters. Then one can estimate the others.

But there is a well-known identification problem: the true value of the impact of constant regressors on the dependent variable is not identified. One can estimate a pseudo-true value, depending on the normalization, not the true value. This has led many researchers to leave out constant regressors. Also the estimates of the normalized FE are typically ignored.

We follow a different route, realizing that these FE contain information on the constant regressors. Our goal is to extract this information, leading to two contributions. The first is that we introduce a constraint on the normalized FE that identifies the true value. As these FE can be estimated, the constraint is testable. This provides an alternative to the leading existing approach by Hausman and Taylor (1981), also denoted by HT, as we will explain.[1]

To illustrate this first contribution, consider a linear country-time panel model. Ignore the constant term for now. The FE is α_i for country i, there is one constant regressor v_i, and ν is the true value of its impact.

One typically leaves out v_i in estimation, that is, normalizes $\nu^0 = 0$. This exemplifies what we call "zero normalization", and we use the superscript 0 to distinguish zero-normalized parameters, which are pseudo-true values, from the true value. An alternative zero normalization is $\alpha_1^0 = 0$. This changes the normalized impact of the

[1] Hausman and Taylor (1981) treat the effect as random instead of fixed. They can of course restrict constant regressors to be uncorrelated with the random effect, a restriction that features many special cases, such as random effects (RE), fixed effects vector decomposition (FEVD) of Plümper and Troeger (2007), and the hybrid approach of Allison (2009, p. 23), which extends Mundlak (1978). But a key contribution is that HT can also allow for correlation, that is, endogeneity of constant regressors. They have an IV procedure where instruments are constructed from within the model by taking averages of variables that also vary over other panel dimensions, such as averages over time of country-time variables in case of country effects. If these instruments are exogenous and relevant, the true value of the impact of the constant regressors is identified. HT can test their moment restrictions, but only if there is overidentification. A potential issue in practice is that the instruments can be weak.

Some papers generalize HT, such as Yang and Schmidt (2020), who consider a three-level "hierarchical" model. Because the underlying ideas are the same, HT remains the relevant reference paper.

constant regressor into $\nu^0 = \nu + \alpha_1/v_1$.[2] Hence, both normalizations pin down their own value for ν^0. Those values generally differ from ν, reflecting that ν is not identified.

Still, both normalizations yield the same sum $\alpha_i^0 + v_i\nu^0$, equal to $\alpha_i + v_i\nu$, so that ν^0 and α_i^0 fully absorb the choice made for the normalization. We can thus safely take some zero normalization and estimate all normalized parameters and the sum $\alpha_i + v_i\nu$.

Now, under the constraint that *all* normalized FE $\alpha_i^0 = 0$, we show that all $\alpha_i = 0$. This implies the true value ν is identified, despite the general notion that this is beyond reach. We also show that this idea generalizes to other normalizations, reflecting that the normalization itself does not identify the true value; it is the constraint.

The constraint may seem restrictive, but it is less restrictive than one may think. After all, it only restricts α_i^0, not the entire sum $\alpha_i^0 + v_i\nu^0$. The sum would be zero if one used the typical normalization $\nu^0 = 0$, but we show how exploiting v_i increases the realism of the constraint. Whether it actually holds depends on the study at hand. This can be tested, because the normalized FE can be estimated.

Notice the fundamental difference between "normalization" and "constraint", which holds throughout the paper. Normalizations are irrelevant for the conditional distribution of the dependent variable (they set parameters in an observationally equivalent way), so in this sense they are without loss of generality. In contrast, we use "constraint" for a restrictive restriction.

The second contribution of the paper concerns the case where our constraint does *not* hold, so there exist nonzero normalized FE. We want to visualize them in a convenient way to extract information on omitted constant regressors. Let us maintain the example above, though no longer ignoring the constant term α, and consider the zero normalization $\alpha^0 = \nu^0 = 0$. This implies that the normalized FE become $\alpha_i^0 = \alpha + \alpha_i + v_i\nu$. Hence, they capture not only the effect of country i, but also the overall intercept and the relevance of v_i, so this overlaps. This blurs the signal from omitted determinants in the estimated α_i^0.

To resolve overlap we introduce "untangling normalization", which makes the normalized FE orthogonal to each other and to constant regressors. Untangling is a normalization, so it is irrelevant for identifying the true value; it is just for interpretation.

Untangling offers several advantages. First, it eases interpretation and is unique. For example, untangling sets 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, easing interpretation. This specific example is not new; see Suits (1984). But we use a much richer 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

[2] The superscript 0 is used for all zero normalizations and its exact meaning appears from the context. The example assumes that the countries are ordered such that $v_1 \neq 0$.

three-dimensional panels. We also give a framework for estimating such FE.

The second advantage of untangling is that the α_i^u capture what is left after accounting for v_i. This facilitates judging how large the deviation from the null hypothesis is, and it helps to find omitted variables, as we will show in our application. Adding newly discovered variables to the model can then shrink the remaining FE.

Our approach is broadly applicable. We illustrate it in a gravity model for exports from OECD countries to the US; see Head and Mayer (2014) for a gravity review. We focus on time FE and have three constant regressors, namely US GDP, world GDP, and the US real effective exchange rate (REER). Our identifying constraint is not rejected, and the constant regressors explain 98% of the time FE, so we leave them out. We thus identify the true values of the constant regressor impacts. This also identifies the US multilateral import resistance. Both results are considered beyond reach in the gravity literature. The REER is a key determinant, accounting for 20% of the time FE, calling for extension of gravity theory, as Klaassen and Teulings (2017) do.

In contrast, if we apply Hausman and Taylor (1981) to our setting, we cannot test their identifying restrictions. If we assume they are valid, we get insignificance for US REER, with a standard error that is over ten times higher than in our method, due to instrument weakness. Hence, our method can be a useful alternative to HT. This is confirmed by the application to pension reforms in Beetsma et al. (2020). In general, our method can help researchers to motivate leaving out FE to get an estimate for constant regressors and, if the constraint is rejected, help them find omitted variables.

The paper is organized as follows. In Section 2 we describe the model and discuss normalization. Section 3 introduces the identifying constraint and shows the differences with Hausman and Taylor (1981). Section 4 sets out untangling normalization. In Section 5 we apply our approach to the gravity model. Section 6 concludes.

2 Model specification

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. This section discusses the model, the perfect multicollinearity involved, and the required normalization, leading to the distinction between pseudo-true and true value of the impact of the constant regressors. This is well known, but we present it in a general way that we exploit later. Identification of the true value is a separate issue and will be addressed in Section 3.

2.1 Model with the full set of FE and constant regressors

The dependent variable is

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

Overparameterization, such as having α and α_i, will be resolved in the next section.

The model has i-, t-, and it-regressors. The vector v_i (w_t) of length K_v (K_w) contains all variables that only vary over countries (time). Hence, v_i and w_t are the constant regressors. Variables that vary over both dimensions are in x_{it} of length K_x. All vectors are column vectors.

To control for some potential omitted regressors, we add unobserved effects. They are grouped in three FE-families. 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 linear trends across countries and consists of τ and τ_i. The θ-family targets the time variation, consisting of θ_t. This adds K_d parameters in total.

The assumptions involving the error term ε_{it} are those that the user prefers, as long as they deliver an estimator of $\alpha + \alpha_i + \ldots + v_i'\nu + w_t'\omega$ in (1), that is, the *sum* of the deterministic and constant regressors parts, not the parts themselves. Hence, our method can be combined with various assumptions and estimators. A typical example is to assume a zero mean of ε_{it} conditional on the regressors in all times, and no cross-sectional correlation, while allowing for heteroscedasticity and serial correlation.

It is convenient to write (1) in matrix form, stacking the time series of the countries:

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

where y and ε stack all y_{it} and ε_{it}, respectively. All deterministic variables constitute the matrix $D = [\iota_N \otimes \iota_T, I_N \otimes \iota_T, \iota_N \otimes [1, 2, \ldots, T]', I_N \otimes [1, 2, \ldots, T]', \iota_N \otimes I_T]$, where ι_n is the vector of ones of length n, and I_n is the identity matrix of order n. The associated FE parameters form $\delta = [\alpha, \alpha_1, \ldots, \alpha_N, \tau, \tau_1, \ldots, \tau_N, \theta_1, \ldots, \theta_T]'$. The constant regressors are stacked into the matrix Z, where its it-th row $[v_i', w_t']$ consists of $K_z = K_v + K_w$ columns, and the corresponding parameter vector is $\gamma = [\nu', \omega']'$. All it-regressors are stacked in X.

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 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 vectors of 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 vectors of 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 $K_d + K_z - m_d - m_z$. Hence, from the sum $D\delta + Z\gamma$ we cannot infer δ and γ.

We thus introduce $m_d + m_z$ normalizations. Their specific forms are left free, to be chosen by the researcher, though they should be linear in the parameters. We call this the general normalization, indicated by g. For later convenience, we formalize it by

$$N^g \begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = 0, \tag{3}$$

where the $(m_d + m_z) \times (K_d + K_z)$ matrix N^g has independent rows, each specifying one normalization. The g-superscript in the general-normalized parameter γ^g makes explicit that it is a pseudo-true value, which generally differs from the true value γ of the impact of the constant regressors, the value one is actually interested in.

A zero normalization is a special case that sets specific elements of δ^0 and γ^0 to zero. Hence, each row of N^0 has a one at the place corresponding to the zero-normalized parameter, and for this matrix, (3) 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.

The total system of equations is the combination of

$$D\delta + Z\gamma = D\delta^g + Z\gamma^g \tag{4}$$

and (3). This gives $K_d + K_z$ independent equations, pinning down δ^g and γ^g.

In the popular special case of zero normalization, (3) sets specific elements of δ^g and γ^g to zero. One then typically continues by removing the corresponding columns of $[D, Z]$ and elements of δ^g and γ^g on the right-hand side of (4) to obtain full column rank. Instead of removing columns and vector elements in (4), we add rows (3) to it. That is equivalent, but our approach is more convenient here, as it can handle many other normalizations in (3), while leaving (4) untouched (Appendix A.2 shows a reduced-columns version of (4) for our approach).

The normalization distributes the sum $D\delta + Z\gamma$ over the fixed effects and the constant regressors. This distribution does not affect other parameters, and the normalization is irrelevant for the conditional distribution of y. It ensures that β, δ^g, and γ^g

can be estimated. Appendix A sets out two estimation methods.

This is where typical fixed-effects modeling stops, motivated by a focus on β. We go further by analyzing δ^g and γ^g in Sections 3 and 4.

3 Identifying the true value of the impact of constant regressors

Our overall approach is split into two parts. Part one concerns normalization, as set out in Section 2.2 above. The normalization is irrelevant for the conditional distribution of the dependent variable y and is only used to obtain an identified model. But the chosen normalization affects γ^g, making its estimate unusable for inference. Put differently, we are interested in the true value, γ, but that has not been identified so far. A simple example is when γ^g is normalized to zero, which says nothing about γ, of course.

Part two of our approach is about a constraint and discussed in Section 3.1 below. There we put a constraint on δ^g to identify γ. This is not limited to some specific normalization, as the g-normalization captures many of them. We focus on the ν-part of γ, that is, the impact of time-constant regressors v_i; the approach for w_t is similar.

3.1 A testable constraint to identify the true value

The identification problem is that an observationally equivalent model results by taking another value instead of ν. That is, taking ν^g instead of ν, defining $\alpha^g + \alpha_i^g$ such that

$$\alpha + \alpha_i + v_i'\nu = \alpha^g + \alpha_i^g + v_i'\nu^g, \tag{5}$$

and then substituting this into (1) gives the same y_{it}. One can thus estimate the pseudo-true values α^g, α_i^g, and ν^g, 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 null hypothesis

$$H_0 : \alpha_i^g = 0 \text{ for all } i. \tag{6}$$

Note that this is not a constraint on the unidentified α_i, but on the normalized α_i^g. The latter can be estimated, so this constraint is testable.

Under the constraint, (5) becomes

$$\alpha_i = \alpha^g - \alpha + v_i' \left(\nu^g - \nu\right), \tag{7}$$

where $\nu^g - \nu$ is unique, because there is no exact linear relationship among the constant regressors. Hence, $[\alpha_1, \ldots, \alpha_N]'$ lies in the column space of $[v_1, \ldots, v_N]'$.

Now, realize that the motivation of including α_i is to control for omitted i-variables, that is, for vectors outside the column space of $[v_1, \ldots, v_N]'$. From (7) we know that such variables are absent. Hence, there is no reason to add α_i, that is, $\alpha_i = 0$ for all i. This implies $\nu = \nu^g$, so the impact of the constant regressor is identified.[3] We thus have a testable constraint to tackle the identification problem. Appendix B presents tests for this constraint, and Section 5 illustrates how they work in practice.

3.2 Relevance of the normalization for the realism of the constraint

Although the chosen normalization g is irrelevant for the conditional distribution of y, and for each g hypothesis (6) identifies ν, the specific choice can matter for the realism of the hypothesis. The reason is as follows, starting from the fact that g consists of $1 + K_v$ normalizations on α^g, α_i^g, and ν^g.

If the normalizations only concern α_i^g, then the null hypothesis constrains its remaining $N - 1 - K_v$ elements. This correctly reflects that the null says the N observations of the sum (5) are driven by a constant and v_i, that is, by $1 + K_v$ determinants.

However, if the normalization is $\alpha^g = 0$ and $\nu^g = 0$, then all explanatory power of the constant term and constant regressors for the sum on the left-hand side of (5) is concentrated into α_i^g, and the null constrains all N of them to zero. This is less realistic than the $N - 1 - K_v$ constraints resulting under the previous normalization, where the explanatory power of the constant term and constant regressors was exploited.

Researchers often normalize at least $\nu^g = 0$. We thus advocate a different approach, one that exploits the information in the constant regressors, so that only part of the sum $\alpha^g + \alpha_i^g + v_i'\nu^g$ ends up in α_i^g. That is, one should normalize some of the α_i^g and leave α^g and ν^g free. In general, for a clean identification analysis, one should normalize parameters that will be constrained by the null hypothesis and leave the remaining involved parameters free. As long as g fulfills this requirement, the specific choice of g does not matter for the realism of the null.[4]

3.3 Comparison to the literature

The leading existing approach to identify ν is due to Hausman and Taylor (1981). HT treat α_i as random and use an IV procedure that depends on the following moment

[3] This resembles the approach in a textbook regression model. That is, consider a cross-section model with one regressor and the standard assumptions. We know that the true value of the impact of the regressor is identified. One could, however, engage into a philosophical exercise by adding an unobserved variable that is an exact multiple of the regressor, and then claim that there is no identification anymore. This is academic, with zero practical relevance. One thus leaves out the unobserved variable and treats the true value as an identified parameter, as we do.

[4] Of course, in case of hypothesis testing, we exclude normalizations that make the null hypothesis impossible, for example, normalizations that fix α_i^g at a nonzero value for some i. Put differently, if the free elements in α_i^g are consistent with the constraint, then the other elements, which result from the free elements and the normalization, must also fulfill the constraint.

restrictions. First, they restrict that a subset of the constant regressors, v_{1i}, are uncorrelated with α_i: $\mathbb{E}\{v_{1i}\alpha_i\} = 0$. Second, consider the other constant regressors, v_{2i}, which are allowed to correlate with α_i and are thus endogenous. HT restrict that the time-averages of a subset of it-regressors, $\bar{x}_{1i}$, are uncorrelated with α_i, so $\mathbb{E}\{\bar{x}_{1i}\alpha_i\} = 0$, and that the $\bar{x}_{1i}$ are relevant instruments for v_{2i}. Let k_1 be the number of instruments and g_2 the number of endogenous constant regressors. Then HT require $k_1 \geq g_2$. All this identifies ν.

If $k_1 > g_2$ (overidentification), HT use the Hausman principle to test their prior restrictions. That is, they compare the estimated β when the two moment restrictions are imposed to the estimate without using them, which is the within-groups estimate.[5]

The difference between our approach and HT is threefold, where we again focus on α_i and v_i. We treat α_i as fixed instead of random. We use another constraint to identify ν. And we can always test the constraint, using estimates of α_i^g instead of β and without requiring overidentification or depending on instrument strength. Let's discuss these in turn. Section 5.4 illustrates them in an empirical application.

First, HT is a hybrid of fixed and random-effects approaches, as the HT subsets $\bar{x}_{2i}$ and v_{2i} are allowed to correlate with α_i, whereas $\bar{x}_{1i}$ and v_{1i} are not. In contrast, we treat α_i as fixed, so that we work in a fully FE setting, allowing all regressors $\bar{x}_i$ and v_i to correlate freely with α_i. This can be valuable in practice, because motivating what regressors in $\bar{x}_i$ and v_i are exogenous can be onerous (Breusch et al. (2011)). Still, if some regressors are known to be exogenous, exploiting that will increase estimation efficiency. Another attractive feature of the FE approach is that it delivers estimates and standard errors of the normalized FE, which helps to see what the model misses.

The second difference concerns how the true value ν is identified. HT use the above moment restrictions for that. Instead, we use a constraint on α_i^g, constraint (6). We can do this because we split the analysis. The first phase handles multicollinearity by some normalization, which is irrelevant for the conditional distribution of the dependent

---

[5] HT encompasses many well-known special cases, including the random effects (RE), Mundlak (1978), and fixed effects vector decomposition (FEVD) approaches. All special cases restrict all constant regressors to be uncorrelated with α_i, that is, $\mathbb{E}\{v_i\alpha_i\} = 0$, and they do not test this. In HT notation, $g_2 = 0$. The special cases differ regarding their restrictions on the correlation between it-regressors and α_i, that is, $\mathbb{E}\{\bar{x}_i\alpha_i\}$, as follows.

The RE approach is the special case that restricts $\mathbb{E}\{\bar{x}_i\alpha_i\} = 0$. Hence $k_1 = K_x$, the number of it-regressors. $k_1 > 0$ enables the familiar Hausman test for random versus fixed effects.

Mundlak (1978) weakens the RE restriction by auxiliary model $\alpha_i = \bar{x}_i^{c\prime}\cdot\pi$ plus noise that has mean zero and is uncorrelated with $\bar{x}_i$, where the c superscript denotes centering and π is the Mundlak parameter. He thus restricts $\mathbb{E}\{\bar{x}_i\alpha_i\} = \mathbb{E}\{\bar{x}_i\bar{x}_i^{c\prime}\}\cdot\pi$. This is a correlated random effects approach, and it is also referred to as the between-within or hybrid approach. As HT set out, it is the special case where $k_1 = 0$. Although Mundlak (1978) has no constant regressors v_i, one can add them, as in the hybrid model of Allison (2009, p. 23).

Finally, FEVD introduced by Plümper and Troeger (2007) leaves $\mathbb{E}\{\bar{x}_i\alpha_i\}$ unrestricted. Hence $k_1 = 0$. This generality also underlies the estimators by Plümper and Troeger (2011), Pesaran and Zhou (2018), and Honoré and Kesina (2017).

variable and generates pseudo-true values that one can estimate, including α_i^g. In the second phase we study the constraint on α_i^g to identify ν. HT do not split, and they impose the identifying restrictions straight away in estimation.

The price we pay is that our constraint (6) can be strong. But this is less of a problem than one may think. If (6) holds (so the two HT moment restrictions are also valid), our approach is more efficient by leaving out the FE, and it avoids potential issues if $\bar{x}_{1i}$ has weak instruments for v_{2i}. If (6) does not hold but the two HT moment restrictions do, we fail to identify ν, but HT can succeed, depending on instrument strength. Still, our FE framework delivers estimated α_i^g with confidence band, which help to find omitted regressors. This has paid off, as in the application of Section 5 the estimated FE reveal the relevance of the real effective exchange rate for explaining exports, giving the idea to extend the theoretical gravity model of exports by this typically omitted determinant. Finally, if neither (6), nor the two HT moment restrictions hold, both approaches fail to identify ν. But, as just described, our estimated α_i^g can still deliver new insights.

The third difference between our approach and HT concerns testing. HT impose the ν-identifying restrictions during estimation. In case of exact identification ($k_1 = g_2$), the identifying restrictions are fulfilled by construction, so that there is no test on their validity. If there are additional exogenous regressors ($k_1 > g_2$), the HT estimate of β becomes different from its within estimate, giving power to the test. As the HT estimate depends on the strength of the instruments $\bar{x}_{1i}$, the test will do as well. In contrast, our split analysis delivers estimates of α_i^g, which always yield a test of the identifying constraint; we do not require overidentification. Moreover, the α_i^g are closer to what one is interested in, the properties of α_i, so we can do more than indirect testing via β. Finally, our test does not depend on instrument strength.

Our method requires enough observations over time to obtain sufficiently accurate estimates of the α_i^g. Here HT may be less demanding, although also there T has to be high enough to get reliable time-averages $\bar{x}_{1i}$ as instruments. We study the relevance of T in Appendix B.4.

4 Untangling normalization

So far, we have estimated the pseudo-true value ν^g, and we have introduced and tested the null hypothesis (6) that identifies the true value ν. Whether this constraint holds is an empirical question. If it does, the FE α_i^g can be left out, and there is no need for normalizing them anymore. However, if the constraint does not hold, one has to choose a normalization.

This section introduces one particularly convenient normalization, untangling normalization. It is denoted by u, so we discuss the special case $g = u$. Importantly, untangling is not used for identifying the true value ν. After all, the identification analysis was for the general normalization g, so that identification of the true value is not driven by a particular normalization — it is due to (6) holding. Moreover, we introduce untangling *after* testing (6), knowing that identification is rejected.

Untangling pins down α_i^u. The key is that it does so in a way that is attractive for interpretation and finding omitted regressors. Hence, the FE framework delivers estimates and covariance matrix of the normalized FE, and untangling enables us to make good use of that benefit.

4.1 The idea and advantages of untangling

The idea of untangling normalization is to handle multicollinearity by making the (normalized) FE orthogonal to each other and to constant regressors, if any. We can now interpret the FE as deviations from both the other FE and the constant regressors. They have been untangled, and each FE-type targets a specific feature of the data.

The main aspects of untangling are as follows. First, consider α^u and α_i^u as an example, so that we need one normalization. Untangling normalization sets the mean of the α_i^u 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. Hence, both effects do not interfere with each other and are assigned to separate parameters, in a unique way. Section 4.2 discusses the details for all FE.

Second, each country-specific regressor in v_i requires one additional normalization. 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. The details are in Section 4.3.

Untangling offers several advantages. It eases interpretation and is unique. Moreover, by focusing on one specific feature of the data and exploiting constant regressors, estimates of α_i^u contribute to finding potentially important omitted regressors. These are advantages over the typically-used normalizations, such as zero normalization, where the overall level and the country deviations are "tangled" into the FE, the normalization depends on ad-hoc choices, and the information in constant regressors is often ignored, as is the case for the $\alpha^0 = \nu^0 = 0$ normalization discussed in the Introduction. The application in Section 5 illustrates these advantages.

4.2 Untangling fixed effects

We first introduce the untangling normalizations concerning the FE themselves. Note that α is the homogeneous type of the intercept fixed effects, so we want α^u to capture the overall intercept, so that we do not normalize it. Likewise, τ^u should capture the overall trend in the model, so we do not normalize that either.

Country-specific effects

To untangle the country FE from the common constant, we normalize the mean of α_i^u to zero, so that they capture the country deviations from α^u. This normalization is well known and goes back to Suits (1984). In formula,

$$\sum_i \alpha_i^u = 0. \tag{8}$$

Country-specific trends

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

$$\sum_i \tau_i^u = 0. \tag{9}$$

Time-specific effects

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

$$\sum_t \theta_t^u \quad = 0 \tag{10}$$

$$\sum_t \theta_t^u \cdot t = 0. \tag{11}$$

4.3 Untangling to exploit constant regressors

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

Country-specific regressors

To clean country FE from the information in the v_i, we make the α_i^u orthogonal to the k-th regressor v_i^k for all k. This gives the following K_v normalizations

$$\sum_i \alpha_i^u v_i^k = 0. \tag{12}$$

This looks like $\mathbb{E}\{v_i \alpha_i\} = 0$, a key relation in the existing literature, which treats α_i as random. It is imposed in existing methods such as RE, FEVD, and to a lesser extent HT. But the difference between (12) and $\mathbb{E}\{v_i \alpha_i\} = 0$ is fundamental. We use (12) just as a normalization, which does not restrict the conditional distribution of the dependent variable, whereas in the existing literature $\mathbb{E}\{v_i \alpha_i\} = 0$ is a moment condition, which is restrictive. Put differently, (12) is to identify the pseudo-true value ν^u, not the true value ν — for the latter we take constraint (6) and that does not rely on untangling. In contrast, $\mathbb{E}\{v_i\alpha_i\} = 0$ is a constraint used to identify the true value ν.

This difference is also visible in estimation and testing. We estimate and test under the g normalization, as explained in Sections 2.2 and 3.1, and these do not rely on (12). On the other hand, existing estimation processes crucially rely on $\mathbb{E}\{v_i\alpha_i\} = 0$, which can then no longer be tested (the exception being HT in case of overidentification).[6]

Time-specific regressors

Similarly, we clean the time FE from the w_t, resulting in K_w normalizations

$$\sum_t \theta_t^u w_t^k = 0. \tag{13}$$

4.4 Untangling in matrix form

Normalizations can be expressed in matrix form, giving a special case of N^g in (3). For untangling, that is, (8)-(13), this is the following matrix N^u. As there are no normalizations on α, τ, ν, and ω, the corresponding columns in N^u are zero.

$$N^u = \begin{array}{c} \begin{matrix} \alpha & \alpha_1 ... \alpha_N & \tau & \tau_1 ... \tau_N & \theta_1 ... \theta_T & \nu' & \omega' \end{matrix} \\ \begin{bmatrix} 0 & 1...1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1...1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1...1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1...T & 0 & 0 \\ 0 & v_1...v_N & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_1...w_T & 0 & 0 \end{bmatrix} \end{array} \quad \begin{array}{l} \text{Row implements:} \\ \sum_i \alpha_i^u = 0 \\ \sum_i \tau_i^u = 0 \\ \sum_t \theta_t^u = 0 \\ \sum_t \theta_t^u \cdot t = 0 \\ \sum_i \alpha_i^u v_i^k = 0 \\ \sum_t \theta_t^u w_t^k = 0. \end{array} \tag{14}$$

5 Application: the gravity model of trade

Many models contain parameters that one thinks are unidentifiable due to added FE. Consider the gravity model of trade.

[6] Interestingly, in a panel with α_i effects only and for the choice $g = u$, our approach encompasses FEVD. After all, in this setting our estimation uses (12) and the sample moment condition for within-groups estimation. The first coincides with the sample version of the moment condition $\mathbb{E}\{v_i\alpha_i\} = 0$ used by FEVD, and the second is the same for FEVD, because also FEVD leaves $\mathbb{E}\{\bar{x}_i\alpha_i\}$ unrestricted. Hence, in this special case our point estimates for v_i and x_{it} will be the same as those of FEVD. Our extension is that we recognize that the estimate for v_i concerns the pseudo-true value ν^u. We then test constraint $\alpha_i^u = 0$ for all i, and if it rejects, we admit that we have no estimate of the true value ν. In contrast, FEVD views the estimate for v_i as an estimate of the true value ν without a test, so by assumption.

5.1 The gravity model and the identification problem

The general idea of the gravity model is that exports from one country to another depend positively on both the exporting and the importing countries' GDPs and negatively on world GDP and the distance between the countries. Distance can be both physical and economic distance, such as trade costs.

Anderson and van Wincoop (2003) show 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 that capture these terms.

If gravity models use bilateral data over time, one could nevertheless control for the multilateral resistances by country-time FE. But then the impacts of country- and time-specific variables, such as exporter, importer, and world GDP, are not identified. This is a well-known problem (Head and Mayer (2014)).

Our method is a potential solution to both problems. We can test whether FE matter after exploiting the information in country- and time-specific observables. If FE do not matter, they can be left out and we have estimated the previously unidentified parameters. That would also mean determination of the economic variables underlying the multilateral resistance terms.

5.2 Model specification

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

$$\begin{aligned} exp_{iUSt} =& \beta_1 gdp_{it} + \beta_2 reer_{it} + \omega_1 gdp_{USt} + \omega_2 gdp_{Wt} + \omega_3 reer_{USt} \\ &+ \alpha + \alpha_i + \tau \cdot t + \tau_i \cdot t + \theta_t + \varepsilon_{it}, \end{aligned} \tag{15}$$

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 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 $reer_{USt}$, so the latter are the constant regressors.[7]

[7] The gdp_{Wt} regressor resembles a Mundlak (1978) term, that is, an average of gdp_{it} over i. Footnote 5 sets out the differences between Mundlak's and our approaches, which we can now illustrate. First, Mundlak uses averages as auxiliary regressors to try to control for the correlation between his random effect and his it-regressors; there are no constant regressors. Instead, we have constant regressors, which are not auxiliary and are motivated by economic theory. Second, gdp_{USt} would be in Mundlak's time effect, creating large correlation with gdp_{it} and its average, invalidating his random effects assumption. In contrast, we treat the time effect as fixed instead of random, thereby recognizing its correlation with

The GDP variables are suggested by the common gravity model. The theory in Klaassen and Teulings (2017) proposes adding real exchange rates, both the bilateral rate rer_{iUSt} and the partners' REERs, here $reer_{it}$ 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 from the gravity literature. Note that it is less standard to include $\tau_i \cdot t$. However, Bun and Klaassen (2007) and Baier et al. (2014) confirm the importance of adding this to account for trends in exports not explained by the regressors, as is often done in the time-series literature.

The model includes lags of all regressors, as traders often entered into contracts in previous periods to export goods in period t, based on export determinants back then. Two lags turn out to be sufficient, and we add them in the form of first differences, for example, Δgdp_{it} and $\Delta gdp_{i,t-1}$. 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, and we ignore them in (15) for simplicity of notation.

The error term ε_{it} has mean zero 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 for strict exogeneity supports this, as leads of regressors have insignificant impacts. The error term is allowed to be heteroscedastic and serially correlated.

We estimate the model using LSDV and the indirect approach of Appendix A.1. This setup is sufficient to illustrate our method. Our main results are robust against different specifications, such as omitting $\tau_i \cdot t$, accounting for non-stationarity and cointegration, and a multiplicative approach estimated by Gamma and Poisson pseudo maximum likelihood (GPML and PPML), following Silva and Tenreyro (2006), as Appendix C shows.

5.3 Data

The data concern $N = 17$ countries, namely the EU-15 countries except for Belgium and Luxembourg, expanded with 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

---

the it- and constant regressors. And we can test instead of assume identification of the true value.

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 is from AMECO, and we use the AMECO exchange rate to express it in national currency. We then divide by the AMECO GDP deflator and 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. Real world GDP in US dollars is 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.

5.4 Identifying the true value

This section illustrates the first contribution of the paper, our test for identification of the true values of the impacts of constant regressors, here gdp_{USt}, gdp_{Wt}, and $reer_{USt}$. The null hypothesis is $\theta_t^g = 0$ for all years t, which is the time equivalent of (6).

As indicated by the g-superscript, we can take any normalization. Following the advice in Section 3.2, and realizing that the model has a constant and trend and that the three t-regressors are included with two lags, we normalize 11 out of the 31 time FE. We also need one normalization on the country FE and one on the country-trend FE. What specific normalization we take is irrelevant for the identification tests.

We now estimate the model and test for identification. We take Wald tests, based on the motivations in Appendix B.1 and the size and power analyses in Monte Carlo Appendix B.4. The diagnostic test that all normalized time FE are zero is 22.77 with 20 degrees of freedom, implying a p-value of 0.30.[8] The sensitivity test is 0.67 with p-value 0.72, which shows that leaving out the time FE does not significantly alter the estimated impacts of gdp_{it} and $reer_{it}$, so this signals no evidence of omitted variable bias. Hence, the tests do not reject identification of the impacts of the three constant regressors. In Section 5.5.2 we will argue that this is most likely not due to a lack of power, as already suggested by the Monte Carlo results in Appendix B.4.

We conclude that three t-regressors have made time FE redundant, so we can safely leave them out. Estimates for the impacts of the t-regressors thus reflect their true values in ω instead of only pseudo-true values. This is remarkable, because identifying such true values has been a notorious problem in the economics literature, not only the literature on gravity models. Of particular interest for the gravity literature is that the redundancy of time FE identifies the multilateral import resistance for the US up to a constant and trend, with $reer_{USt}$ as key determinant. This also suggests that $reer_{it}$ is an important determinant of the multilateral export resistance of country i. The

[8] The standardized Wald test, discussed in footnote 16, yields a similar p-value of 0.33.

rightmost column in Table 1 shows the resulting estimates. All signs can be explained within the theoretical gravity model.[9]

Table 1: Estimation results for exp_{iUSt} based on model (15)

	Without $reers$			With both $reers$		
Specification	1	2	3	4	5	6
Included	θ_t	θ_t	θ_t	θ_t	θ_t,DL	DL
Normalization	0	0,u0	u	u	u	–
Figure of FE	1a	1b,1c	1d=2a	2b	2c	–
gdp_{it}	0.60 *	0.60 *	0.60 *	0.80 *	1.12 *	1.13 *
	(0.23)	(0.23)	(0.23)	(0.24)	(0.25)	(0.24)
$reer_{it}$				0.42 *	0.66 *	0.68 *
				(0.12)	(0.15)	(0.14)
gdp_{USt}			3.43 *	3.03 *	2.35 *	2.30 *
			(0.30)	(0.29)	(0.39)	(0.37)
gdp_{Wt}			−3.59 *	−1.84 *	−1.18	−1.23
			(0.41)	(0.45)	(0.73)	(0.72)
$reer_{USt}$				−1.02 *	−1.02 *	−1.01 *
				(0.10)	(0.13)	(0.13)
Wald tests						
$\theta_t^g = 0$	770.04 *	545.14 *	222.57 *	85.31 *	22.77	–
	[0.00]	[0.00]	[0.00]	[0.00]	[0.30]	
$\beta = \beta\|_{\theta_t^g=0}$	2.50	80.69 *	1.34	9.93 *	0.67	–
	[0.11]	[0.00]	[0.25]	[0.01]	[0.72]	
R_θ^2	0	0	0.65	0.91	0.98	–

Static models have 561 observations and distributed lag (DL) models 527. DL models have two lags of each regressor; we display the long-run effects. The normalizations used in specifications 1 and 2 are made explicit in the corresponding FE figures.
The first Wald is the diagnostic test motivated in Appendix B.2 and tests 33, 31, 29, 28, 20 independent constraints for specifications 1-5. The second Wald, labeled $\beta = \beta|_{\theta_t^g=0}$, is the sensitivity test of Appendix B.3, showing how sensitive the estimator for β of the level variables is to setting $\theta_t^g = 0$, which concerns 1, 1, 1, 2, 2 constraints.
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 we use throughout the paper.

[9]The literature proposes the Hausman and Taylor (1981) method, set out in Section 3.3. Let's apply it here and show that our method can be a useful alternative. Hence, we now treat θ_t as random (α_i and τ_i are still fixed). If gdp_{USt} is endogenous regarding θ_t, business cycle co-movement across countries makes it likely that also gdp_{it} and thus the cross-sectional average $\overline{gdp}_t$ and gdp_{Wt} are endogenous. This gives two endogenous constant regressors and only $\overline{reer}_t$ is a potential instrument. Therefore, to avoid violating the order condition, we must treat all GDP regressors as exogenous. Next, suppose we allow $reer_{USt}$ to be endogenous. Because the US is an important trading partner of all countries i, there will be a negative correlation between $reer_{USt}$ and $reer_{it}$, so that $\overline{reer}_t$ is also endogenous. We get an exactly identified HT model, where $\overline{gdp}_t$ is an instrument for $reer_{USt}$. Hence we cannot use the HT test for exogeneity. But our non-rejection of $\theta_t^g = 0$ suggests that the HT moment restrictions are reasonable here. The Stata estimates and standard errors for the constant regressors are: for gdp_{USt} 2.30 (0.57), for gdp_{Wt} 1.39 (4.12), and for $reer_{USt}$ -2.25 (1.62). The latter two estimates differ substantially from ours and the standard errors are much larger. The reason is that $\overline{gdp}_t$ is a weak instrument for $reer_{USt}$. Hence, even when the HT moment restrictions are reasonable, our method can be a useful alternative.

5.5 Untangling

The second contribution of the paper addresses the question what to do when the null hypothesis $\theta_t^g = 0$ is rejected. We create such a situation within the framework just used.[10] More specifically, we start from the empirical gravity model that is now common practice (Head and Mayer (2014)), which is (15) without the REERs ($\beta_2 = \omega_3 = 0$). Note that the latter is a constraint, no normalization. There are no lagged regressors.[11]

We thus deliberately take a step back by reducing the explanatory part of the model. This allows us to illustrate how an empirical researcher, starting from a well-known model, can use untangling normalization of Section 4 for better interpretation and thereby improve the specification. At the same time, we have untangling *after* the test outcome of not rejecting identification in Section 5.4, which highlights that the test in no way depends on untangling normalization, a key point to be realized.

5.5.1 Untangling helps interpretation

The first advantage of untangling is that it facilitates interpretation. This section illustrates that by varying the normalization, while keeping the rest of the model the same. Hence, there is one estimation, and then the estimates are simply transformed to fulfill other normalizations. All normalizations have one on the country FE and one on the country-trend FE, but we focus on the remaining four choices, two due to the time FE and two regarding the constant regressors gdp_{USt} and gdp_{Wt}.

One typically chooses a zero normalization. We study two such choices, both having $\omega^0 = 0$. Figure 1a normalizes $\alpha^0 = \tau^0 = 0$ and shows the estimated normalized time FE θ_t^0 and confidence band. Their mean is nonzero, reflecting that they are affected by the overall means of the dependent and explanatory variables. They exhibit some variation over time, but this seems a minor feature.

Figure 1b normalizes $\theta_{T-1}^0 = \theta_T^0 = 0$, giving a positive mean and a shrinking confidence band. The changes compared to Figure 1a exemplify the well-known impact of different zero normalizations on the FE estimates, which hampers their interpretation.

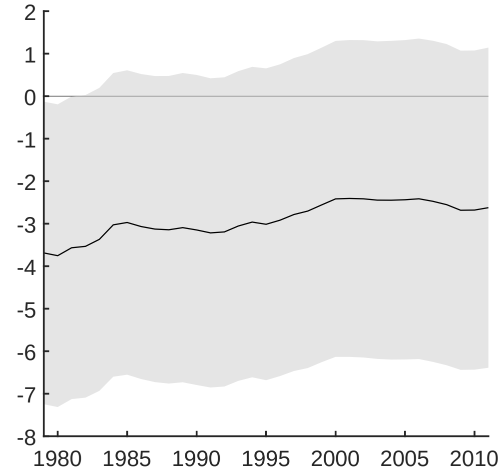
It is not yet clear how important the variation over time is, as that may be dominated by the constant or the trend. To have this curvature information visible right away, it would have been appealing to let the normalization split off the level and trend information from the time FE. That is what untangling normalization does, by (10) and (11), in a unique way.

Figure 1c shows the results, where for the moment we do not yet exploit the constant

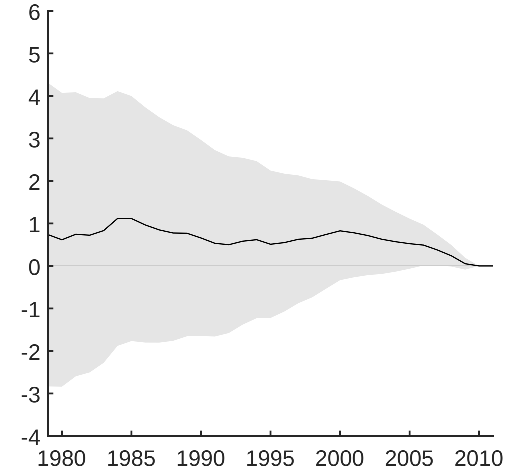
[10] Another possibility is to take other data. Appendix C exemplifies that by adding pre-1979 data.

[11] Throughout this section we consider the number of countries $N = 17$ to be sufficiently large to rely on asymptotic normality when discussing the estimated time FE, stimulated by our own simulation results and those in the literature as reported in Appendix B.4.

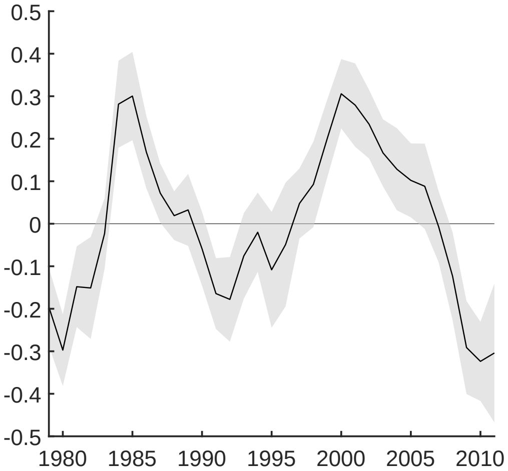
Figure 1: From zero-normalized to untangled time FE in the common gravity model



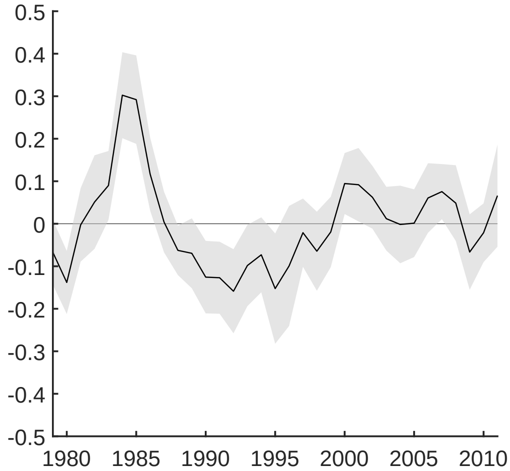
(a) θ_t^0 if $\alpha^0 = \tau^0 = 0$ and $\omega^0 = 0$



(b) θ_t^0 if $\theta_{T-1}^0 = \theta_T^0 = 0$ and $\omega^0 = 0$



(c) Untangled θ_t^{u0}, but still $\omega^{u0} = 0$



(d) Untangled θ_t^u

All bands around the time FE in the paper are 95% simultaneous confidence bands. They are sup-t bands, following the suggestion by Montiel Olea and Plagborg-Møller (2019). The sup-t band is the narrowest band in the class of confidence bands that scale up the pointwise band (which simply combines the confidence intervals of the underlying parameters) by one parameter to guarantee the required simultaneous coverage. That parameter is about 1.0 for the bands based on zero normalization and 1.6 based on untangling.

regressors. Hence, this is a combination of untangling and zero normalization, indicated by the superscript $u0$. These FE are more informative and easier to interpret than the zero-normalized ones. Note the economic downturn in the early 1990s, the dot-com bubble, and the recent financial crisis. Footnote 13 will give a more complete analysis of the correspondence between untangled FE and the business cycle.

In addition, untangling has resulted in a more informative confidence band. Untangling can thus better capitalize on an advantage of a fixed instead of random effects approach, which is that the former delivers insights into the accuracy of all estimated time effects. Overall, the estimated θ_t^{u0} indicate the model misses export determinants.

Figure 1d shows the FE from untangling normalization, so we now exploit the constant regressors gdp_{USt} and gdp_{Wt}, using (13). Untangling shrinks and cleans the FE, so it better shows when exports deviate from what the common gravity model explains.

This section has only varied the normalization. Section 3.2 has shown that normalizations can matter for the restrictiveness of our identifying constraint $\theta_t^g = 0$ for all t. Hence, as a side issue, we now calculate our Wald tests for the normalizations in Figures 1a-1d. The bottom parts of columns 1-3 in Table 1 show the results.

The drop in the diagnostic Wald from Figure 1a to 1b (column 1 to 2) reflects that leaving the overall constant and trend free makes the constraint more realistic, confirming our advice in Section 3.2 to normalize parameters that also appear in the null hypothesis, here θ_t^g.[12] How the θ_t^g are normalized does not matter, which is illustrated by the equal Walds for Figures 1b and 1c.

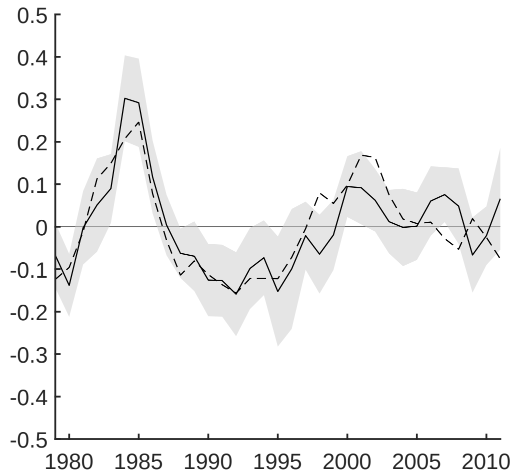
Moving to Figure 1d and thus exploiting the constant regressors reduces the Walds. This is again in line with our advice and reflects that the constraint concerns the FE left over after accounting for constant regressors. In sum, untangling normalization, or any other approach with the normalizations on θ_t^g only, provides the cleanest indication of how close one is to identification.

5.5.2 Untangling helps to find omitted variables

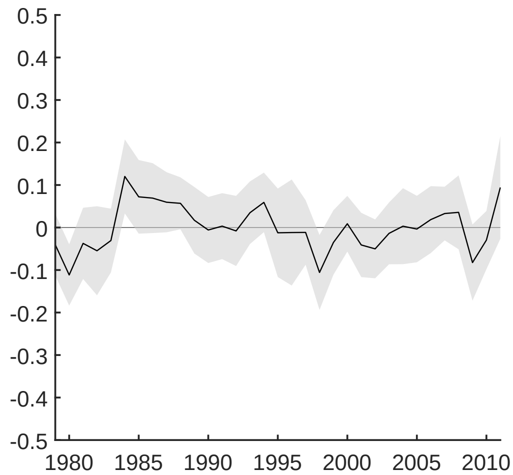
Figure 2a replicates 1d but adds minus US REER as a dashed line (after detrending, demeaning, and scaling). The resemblance with the FE is striking. For example, consider the eighties, where the dollar bubble stimulates and then hampered exports to the US. Hence, the US REER seems important in explaining time FE, illustrating the second advantage of untangling, that it can reveal omitted regressors. Klaassen and Teulings (2017) thus extend gravity theory and find that $reer_{it}$ and $reer_{USt}$ matter. Hence, here untangling not only reveals t-regressors, but also indirectly an it-regressor.

[12] Our advice also avoids chance dependence of the sensitivity Wald. That is, the small value for Figure 1a is because the additional constant and trend restrictions make that the estimated β_1 and thus the test depend on the coincidental level and scale of the variables. Our advice avoids this.

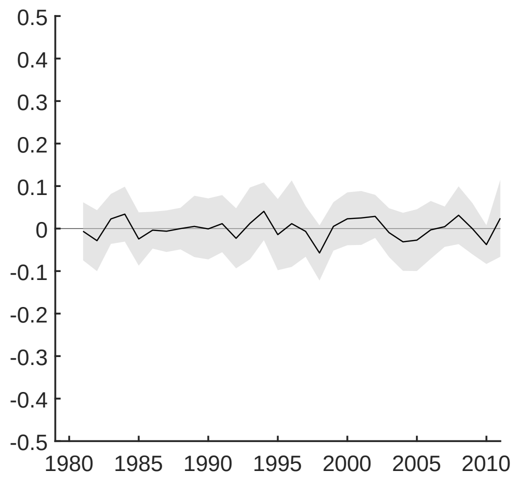
Figure 2: Untangled time FE θ_t^u when adding $reer$s to the common gravity model



(a) Common gravity model 3; dashed $-reer_{USt}$



(b) Gravity model with $reer$s: model 4



(c) Gravity model with $reer$s and lags: model 5

We thus add $reer_{it}$ and $reer_{USt}$ to the model, that is, we leave β_2 and ω_3 free; see column 4 in Table 1.[13] Compared to column 3, we observe a notable change in the estimated impact of gdp_{Wt}. This indicates that adding the REERs substantially mitigates omitted variable bias.

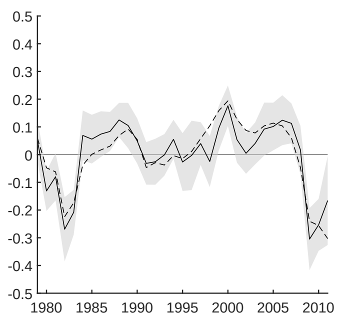
Figure 2b presents the estimated FE θ_t^u. There is again a large reduction of the FE, and they are now close to zero. Taken together, adding three t-regressors has reduced the FE substantially, from those in Figure 1c to Figure 2b. We can explain almost all of the $T = 33$ time FE by just three variables. This can be quantified by $R_\theta^2 = 91\%$, which is defined in the note to Table 1.

Finally, we add two lags of all regressors, in first-difference form, as in Section 5.2. The lags of it-regressors will take some noise out of θ_t^u, and lagged t-regressors 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 and find omitted export determinants, in agreement with the idea of untangling. We thus again take a distributed lag (DL) model and still allow for unrestricted serial correlation in the error. The estimates are in Table 1 specification 5. Figure 2c shows the estimated θ_t^u. They are close to zero, and $R_\theta^2 = 98\%$.

We have now arrived at the same model as used in Section 5.4 for testing the constraint that identifies the true value ω. But we now have untangling normalization. This normalization falls in the class of g-normalizations advised by Section 3.2, that is, where the normalizations are on the time FE. Hence, it is not surprising that recalculating the Wald tests for untangling yields exactly the same values, confirming that identification does not depend on untangling. It is the constant regressors that have led to the non-rejection; they shrink the θ_t^u.

Also note that the confidence band in Figure 2c includes zero for all t. This signals that the insignificance of the Wald tests is not due to a lack of power caused by aggregating the information on all FE into a single statistic. In fact, the band in Figure 2b, which is already close to zero, comes with clear rejections of both Wald tests, namely 85.31 (p-value 0.00) and 9.93 (0.01). This indicates that the tests have power, in line with the Monte Carlo results in Appendix B.4.

[13] An alternative way to show that untangling can help researchers to find omitted regressors is by taking this more general model 4 and then leaving out the t-regressors one by one. Leaving out US GDP gives the estimated untangled FE in the solid line in the figure on the right, and the dashed line is the omitted regressor itself. There is a strong resemblance. This again shows that untangling can reveal omitted regressors. If we redo this for world GDP, the resemblance is much weaker (figure not reported), but for US REER it is strong (the figure is not reported but is almost equal to Figure 2a).



6 Conclusion

We have shown that in fixed-effect models the true values of the impacts of constant regressors are identified if normalized FE are zero, whatever the normalization. This yields a novel approach that, if the constraint holds, resolves a notorious problem in the literature. The constraint is testable. Our approach can help researchers to motivate leaving out FE. It is an alternative to Hausman and Taylor (1981), and both methods have their own merits.

We have applied our method to a panel gravity model for exports to the US. With only three t-regressors — US GDP, world GDP and US REER — the year FE become redundant, so that we have identified the true values of their impacts, even though that is typically considered beyond reach. This also means that we have identified 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.

Our second contribution concerns the case where the constraint does not hold. For that, we have introduced untangling normalization. It disentangles FE-types from each other and from constant regressors, which eases the interpretation of the normalized FE. This also helps researchers to find omitted variables.

The gravity application has illustrated how untangling can visualize the information in estimated normalized FE, and how untangling has revealed the relevance of three t-regressors. The business cycle pattern in exports is well known. But untangling has also revealed the importance of the US REER. We thus recommend giving exchange rates a more prominent role in gravity theory.

This paper has used parameter homogeneity to simplify the exposition, and it has turned out to be sufficient. In future work, one may want to allow for heterogeneity, further shrink the FE, and make our identifying constraint more realistic. Moreover, as the generalization to ijt-panels is straightforward, untangling can facilitate studies of financial or trade relations involving many sectors j. Finally, untangling illustrates the value of information in FE, which may stimulate further research on their estimation.

References

Allison, P. (2009): *Fixed Effects Regression Models*, SAGE Publications, Inc.

Anderson, J. E. and E. van Wincoop (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *The American Economic Review*, 93, 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, 339–350.

Basu, D. (2020): "Bias of OLS Estimators due to Exclusion of Relevant Variables and

Inclusion of Irrelevant Variables," *Oxford Bulletin of Economics and Statistics*, 82, 209–234.

Beetsma, R., F. Klaassen, W. Romp, and R. van Maurik (2020): "What drives pension reforms in the OECD?" *Economic Policy*, xx, xxx–xxx.

Breusch, T., M. B. Ward, H. T. M. Nguyen, and T. Kompas (2011): "On the Fixed-Effects Vector Decomposition," *Political Analysis*, 19, 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, 473–496.

Donald, S., G. Imbens, and W. Newey (2003): "Empirical likelihood estimation and consistent tests with conditional moment restrictions," *Journal of Econometrics*, 117, 55–93.

Evans, G. B. A. and N. E. Savin (1982): "Conflict among the Criteria Revisited; The W, LR and LM Tests," *Econometrica*, 50, 737–748.

Fernández-Val, I. and M. Weidner (2018): "Fixed Effects Estimation of Large-T Panel Data Models," *Annual Review of Economics*, 10, 109–138.

Ghysels, E., J. B. Hill, and K. Motegi (2020): "Testing a large set of zero restrictions in regression models, with an application to mixed frequency Granger causality," *Journal of Econometrics*, 218, 633 – 654.

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

Head, K. and T. Mayer (2014): *Gravity Equations: Workhorse,Toolkit, and Cookbook*, Elsevier, vol. 4 of *Handbook of International Economics*, chap. 3, 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, 543–558.

Kiviet, J. F. (2012): "Monte Carlo Simulation for Econometricians," *Foundations and Trends(R) in Econometrics*, 5, 1–181.

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

Lu, X. and L. Su (2020): "Determining individual or time effects in panel data models," *Journal of Econometrics*, 215, 60–83.

Magnus, J. R. and A. L. Vasnev (2007): "Local sensitivity and diagnostic tests," *Econometrics Journal*, 10, 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, 655–680.

Montiel Olea, J. L. and M. Plagborg-Møller (2019): "Simultaneous confidence bands: Theory, implementation, and an application to SVARs," *Journal of Applied Econometrics*, 34, 1–17.

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

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

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

Pesaran, M. H. and Q. Zhou (2018): "Estimation of time-invariant effects in static panel data models," *Econometric Reviews*, 37, 1137–1171.

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, 124–139.

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

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

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

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

Yang, Y. and P. Schmidt (2020): "An econometric approach to the estimation of multi-level models," *Journal of Econometrics*, xx, xxx–xxx.

Appendix

A Estimation of normalized parameters

It is important to realize that this appendix (on estimation) and the next one (on testing) use the general setting of the g-normalization. Hence, both do *not* impose the untangling normalization of Section 4. The latter is just a special case that results by taking (14) for N^g, which happens to simplify the formulas.

Consider mean equation (2). Normalizations only affect the lower-dimensional part of the regressors, so the focus in this appendix is on $D\delta + Z\gamma$. We introduce two methods of estimation, an indirect and a direct approach.

A.1 Indirect estimation: renormalizing zero-normalized estimates

The indirect estimation method consists of two steps. It first estimates (2) using a zero normalization and then renormalizes the estimates into the desired g normalization.

A.1.1 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 dummies and regressors 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.

A.1.2 Renormalizing the estimates

The second step renormalizes δ^0 and γ^0 into the general-normalized parameters δ^g and γ^g, that is, it redistributes the sum $D\delta^0 + Z\gamma^0$ over $D\delta^g$ and $Z\gamma^g$. For both normalizations, system (3)-(4) holds, so we obtain

$$\begin{bmatrix} D & Z \\ & N^g \end{bmatrix} \begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = \begin{bmatrix} D & Z \\ & N^0 \end{bmatrix} \begin{bmatrix} \delta^0 \\ \gamma^0 \end{bmatrix}. \tag{A.1}$$

To solve for δ^g and γ^g, define

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

and pre-multiply (A.1) with $R^{g\prime}$. As R^g has full column rank, we obtain

$$\begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = R^{0g} \begin{bmatrix} \delta^0 \\ \gamma^0 \end{bmatrix}, \tag{A.3}$$

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, and the matrix only consists of observables.[14]

Hence, to obtain the estimates and estimated variance for the g-normalized parameters, we take both for the zero-normalized parameters and apply (A.3). Therefore, no additional estimation or standard error correction is needed.

A.2 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 Appendix B.

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,

$$N^g P = \begin{bmatrix} N_f^g & N_r^g \end{bmatrix} \text{ and } P' \begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = \begin{bmatrix} \delta_f^g \\ \gamma_f^g \\ \delta_r^g \\ \gamma_r^g \end{bmatrix}. \tag{A.4}$$

Hence, 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 (3), writing N^g as $N^g P P'$, and using (A.4) gives

$$\begin{bmatrix} \delta_r^g \\ \gamma_r^g \end{bmatrix} = -N_r^{g-1} N_f^g \begin{bmatrix} \delta_f^g \\ \gamma_f^g \end{bmatrix}. \tag{A.5}$$

[14] Computing R^{0g} depends on multiplications involving R^g and R^0, which have many rows. This can be simplified. 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 (A.1), 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 (A.3). This yields the same g-normalized parameters.

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

$$\begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = F^g \begin{bmatrix} \delta_f^g \\ \gamma_f^g \end{bmatrix}, \tag{A.6}$$

where

$$F^g = P \begin{bmatrix} I_{K_d+K_z-m_d-m_z} \\ -N_r^{g-1} N_f^g \end{bmatrix}. \tag{A.7}$$

We partition F^g into four blocks $[F_{11}^g\ F_{12}^g;\ F_{21}^g\ F_{22}^g]$ such that (A.6) yields

$$D\delta^g + Z\gamma^g = D^g \delta_f^g + Z^g \gamma_f^g, \tag{A.8}$$

where $D^g = DF_{11}^g + ZF_{21}^g$ and $Z^g = DF_{12}^g + ZF_{22}^g$. We have thus incorporated the g-normalization into the regressor matrix. Hence, we are in a standard setting, where δ_f^g and γ_f^g can be estimated and (A.6) then gives the estimate of the full vector.[15]

Results (A.7) and (A.8) simplify if the normalization does not involve γ^g, that is, if the rightmost K_z columns in N^g are zero. This holds for the special normalizations advocated in Sections 3.2 and 4, among other ones. In this paragraph we thus consider $\gamma^g = \gamma_f^g$. Then (A.6) implies $[F_{21}^g, F_{22}^g] = [0, I_{K_z}]$. Moreover, the rightmost K_z columns in N_f^g, which refer to γ^g by construction of P, contain only zeros. Then the same holds for $N_r^{g-1} N_f^g$. Hence, considering the complete F^g matrix, (A.7) implies that its rightmost K_z columns consist of zeros except for a block I_{K_z}. The P matrix in (A.7) permutes the rows such that I_{K_z} ends up at the rows corresponding to the elements of γ^g, that is, the bottom rows. Hence, above those rows, the rightmost K_z columns in F^g contain only zeros. Hence, F^g is block diagonal:

$$F^g = \begin{bmatrix} F_{11}^g & 0 \\ 0 & I_{K_z} \end{bmatrix}. \tag{A.9}$$

As a result, $D^g = DF_{11}^g$ and $Z^g = Z$. Hence, Z is no longer transformed, reflecting there is no normalization on γ^g.

B 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^g, leaving γ unidentified. But $\alpha_i^g = 0$ for all i implies $\nu^g = \nu$, and $\theta_t^g = 0$ for all t implies $\omega^g = \omega$, as explained in Section 3, so that we have constraints that identify (parts of) $\gamma = [\nu', \omega']'$.

[15]To see that $[D^g, Z^g]$ 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 (A.8) implies the same for $[D^g, Z^g]$. Because this dimension equals the number of columns of the latter matrix, its columns are independent.

This appendix introduces two approaches to test such constraints. We discuss them for the general normalization, and untangling normalization of Section 4 again follows as a simple special case. The first approach 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 approaches are special cases of the following more general testing procedure.

B.1 Testing under normalizations

As long as parameters are not restricted in normalizations, such as ν^g if restrictions are on α_i^g instead, we can use a standard way to test constraints, such as a Wald test. But parameters that are restricted 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

$$H_0 : C \begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = c, \tag{B.1}$$

where C is the constraint matrix with independent rows, and c is a vector of constraint values. Testing this involves two problems. First, the estimator of $[\delta^{g\prime}, \gamma^{g\prime}]'$ has a singular covariance matrix, due to the normalization. This is resolved by substituting (A.6) into (B.1), giving $CF^g \left[\delta_f^{g\prime}, \gamma_f^{g\prime}\right]' = c$, where the vector of free parameters can be estimated in the standard way with a non-singular covariance matrix (see Appendix A.2).

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

$$H_0 : C^g \begin{bmatrix} \delta_f^g \\ \gamma_f^g \end{bmatrix} = c^g. \tag{B.2}$$

One could take the Wald test using the χ^2_Q-distribution, where Q is the number of independent constraints.[16] It may also be informative to study the constraints individually,

[16] Q may grow with the sample size. For example, if the null hypothesis is that all $\alpha_i^g = 0$, and one uses an approximation based on $N \to \infty$, then the degrees of freedom of χ^2_Q diverge. Donald et al. (2003) show that even then the χ^2_Q-distribution applies. Because such an approach is correct for fixed Q as well, they prefer it over a standardized test, such as the standardized Wald statistic $(\text{Wald} - Q)/\sqrt{2Q}$ with a standard normal asymptotic distribution. Lu and Su (2020) test for the presence of FE when N and T pass to infinity simultaneously. Ghysels et al. (2020) introduce a test for many zero restrictions, but a Wald test is preferable here because of higher power, as discussed in Appendix B.4.

for example, by t-tests.

B.2 Diagnostic test

The g-normalized FE α_i^g and θ_t^g represent i-variables and t-variables that are relevant for y_{it} but omitted from the model. Our first test is about existence of such omitted variables, so the null hypothesis constrains α_i^g and/or θ_t^g to zero. This means that part of δ^g is constrained, and we denote that part by a subscript 0. Hence, the null is $\delta_0^g = 0$. This is a special case of (B.1). Hence, defining C accordingly, setting $c = 0$, and following the approach of the previous section gives a test, such as the Wald test. This is our diagnostic test.

The constraint is sufficient for identification of ν and/or ω. It is not necessary, because α_i^g and/or θ_t^g 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.

B.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 some parameters. This is similar to the idea underlying an omitted variables bias test. We compare the unconstrained estimator $\widehat{\beta}$ to the estimator $\widetilde{\beta}$ under the constraint $\delta_0^g = 0$, which is equivalent to setting its free part $\delta_{f0}^g = 0$. If $\widehat{\beta} - \widetilde{\beta} = 0$, the estimator of β is insensitive to the FE, which supports leaving them out.

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

$$\begin{bmatrix} \widehat{\delta}_{f\emptyset}^g \\ \widehat{\gamma}_f^g \\ \widehat{\beta} \end{bmatrix} - \begin{bmatrix} \widetilde{\delta}_{f\emptyset}^g \\ \widetilde{\gamma}_f^g \\ \widetilde{\beta} \end{bmatrix} = -\Big(\big[D_\emptyset^g, Z^g, X\big]' \big[D_\emptyset^g, Z^g, X\big] \Big)^{-1} \big[D_\emptyset^g, Z^g, X\big]' D_0^g \cdot \widehat{\delta}_{f0}^g, \quad \text{(B.3)}$$

where $\delta_{f\emptyset}^g$ collects the elements of δ_f^g that are not in δ_{f0}^g, and $D_\emptyset^g$ and D_0^g are the corresponding submatrices of D^g defined below (A.8).[17] We know the distribution of $\widehat{\delta}_{f0}^g$ and thereby of $\widehat{\beta} - \widetilde{\beta}$. We can thus test whether its realization differs significantly from zero, for example by a Wald test. This is our sensitivity test. It essentially

[17] To derive (B.3), combine model equation (2) with (4) and (A.8). Then split off δ_{f0}^g by partitioning the regressors into $L = \big[D_\emptyset^g, Z^g, X\big]$ and D_0^g, and collecting the parameters for L by $\lambda = [\delta_{f\emptyset}^{g\prime}, \gamma_f^{g\prime}, \beta']'$. Now the model becomes $y = L\lambda + D_0^g \delta_{f0}^g + \varepsilon$. Then $[\widehat{\lambda}', \widehat{\delta}_{f0}^{g\prime}]'$ depends on $([L, D_0^g]'[L, D_0^g])^{-1}$. Applying the partitioned inverse formula to the latter, and substituting the resulting expression for $\widehat{\delta}_{f0}^g$ into that for $\widehat{\lambda}$ yields (B.3).

takes (B.2) and uses a specific linear combination of δ^g_{f0}, illustrating that not the mere absence of omitted variables ($\delta^g_{f0} = 0$) is crucial, but rather how much a combination of them matters for estimating parameters of interest.

To interpret (B.3), distinguish two parts on the right. At the end, we have the diagnostic part, $\widehat{\delta}^g_{f0}$, 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 $\widehat{\delta}^g_{f0}$ 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.

B.4 Monte Carlo study

This appendix presents a concise Monte Carlo analysis of the above diagnostic and sensitivity tests. Given their popularity, we focus on Wald tests. We only consider FE α_i, so the null hypothesis is that all normalized country fixed effects are zero; as in (6). The results, however, are one-to-one applicable to a setting with only time FE θ_t.[18]

For some designs, exact finite-sample results of the Wald tests exist. Still, practitioners typically use the χ^2-approximation, which yields oversize for small T, so one relevant question is how quickly that disappears when T grows in a panel. Moreover, what happens if we let N and thus the number of fixed effects grow large? How powerful are the Wald tests for our hypothesis? How do correlations between fixed effects and observed regressors matter? Together with the empirical application in Section 5, the Monte Carlo answers to these questions will illustrate the potential of our approach.

B.4.1 Design

The model contains country FE, one constant regressor v_i, and one it-regressor x_{it},

$$y_{it} = \alpha^0 + \alpha^0_i + \nu^0 v_i + \beta x_{it} + \varepsilon_{it}, \tag{B.4}$$

[18] We focus on testing, and under the null hypothesis there are no fixed effects α^g_i to be estimated. Still, let us briefly address the estimated FE if the null does not hold, in particular (small-sample) bias. For notational convenience, do as if $\nu^g = 0$. If the regressors are strictly exogenous regarding the error, LSDV is an unbiased estimator of the α^g_i, irrespective of T. If, in addition, the error is normally distributed, the estimated α^g_i are also normal.

For alternative error assumptions, consider Fernández-Val and Weidner (2018), 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^g_i this means $1/T$. Buddelmeyer et al. (2008) use simulations to study the bias and for their settings the biases are fairly small for $T = 20$. Both papers also show how bias correction can further improve small-sample properties.

where we have normalized the FE. The exact normalization is irrelevant in the Monte Carlo analysis, but a concrete example is $\alpha^0_{N-1} = \alpha^0_N = 0$, so that the null hypothesis is $\alpha^0_1 = \ldots = \alpha^0_{N-2} = 0$. We estimate by LSDV and study degrees-of-freedom-corrected Wald tests, using 5% χ^2-critical values.

The data generating process (DGP) allows for correlation between α_i, v_i, and x_{it}, in line with our fixed effects framework. More specifically, we generate

$$y_{it} = \alpha_i + \varepsilon^y_{it} \tag{B.5}$$

where

$$\begin{aligned} \alpha_i &= \alpha_v \varepsilon^v_i + \alpha_x \varepsilon^x_i + \varepsilon^\alpha_i && \text{(B.6)} \\ v_i &= \nu_x \varepsilon^x_i + \varepsilon^v_i && \text{(B.7)} \\ x_{it} &= \varepsilon^x_i + \varepsilon^\chi_{it}. && \text{(B.8)} \end{aligned}$$

We leave out a constant, v_i, and x_{it} in (B.5), because they would not affect our Wald tests given the above model. The parameters $\alpha_v, \alpha_x, \nu_x$ govern the key correlations. The innovations $\varepsilon^y_{it}, \varepsilon^v_i, \varepsilon^x_i, \varepsilon^\chi_{it}, \varepsilon^\alpha_i$ are iid with zero means. Without loss of generality in our exercise, we fix the variances of $\varepsilon^y_{it}, \varepsilon^v_i, \varepsilon^x_i$ at unity.[19] The variance of ε^χ_{it} is $\sigma^2_\chi > 0$, and that of ε^α_i is $\sigma^2_\alpha \geq 0$. All innovations are normally distributed, though we also address the uniform distribution.

Instead of choosing the five DGP parameter values directly, we first move to a more orthogonal parameter space, as in Kiviet (2012). The five base parameters follow in a straightforward manner. First, $\sigma^2_\chi > 0$ is a base parameter by itself and quantifies the variation of x_{it} within a country over time relative to the pure cross-sectional variation due to ε^x_i. Second, the variance $\text{Var}(\alpha_i) = \alpha^2_v + \alpha^2_x + \sigma^2_\alpha \geq 0$ is the total variation of the country effect. Finally, consider the correlations between α_i, v_i, and ε^x_i, that is, $\rho_{\alpha v} \in [-1, 1]$, $\rho_{\alpha x} \in (-1, 1)$, and $\rho_{vx} \in (-1, 1)$, where the first two are only defined if $\text{Var}(\alpha_i) > 0$.[20] The correlation between α_i and v_i consists of a direct and an indirect part via ε^x_i, and we simply define the direct part $\rho^d_{\alpha v}$ by splitting off the indirect part, so $\rho^d_{\alpha v} = \rho_{\alpha v} - \rho_{\alpha x}\rho_{vx}$ (one could also take the partial correlation). We take $\rho^d_{\alpha v}$, $\rho_{\alpha x}$, and ρ_{vx} as base parameters.

Below, we choose base parameters within the above ranges. We then derive the

[19]The Wald tests are invariant to these fixations. First, multiplying all five innovation variances by some positive factor does not change the Wald tests, so we can set that of ε^y_{it} to unity. Moreover, multiplying ε^v_i by a nonzero factor is absorbed by dividing ν^0 by that factor, and similarly for ε^x_i, ε^χ_{it}, and β, without changing the Wald tests, so that we can fix the variances of ε^v_i and ε^x_i at unity.

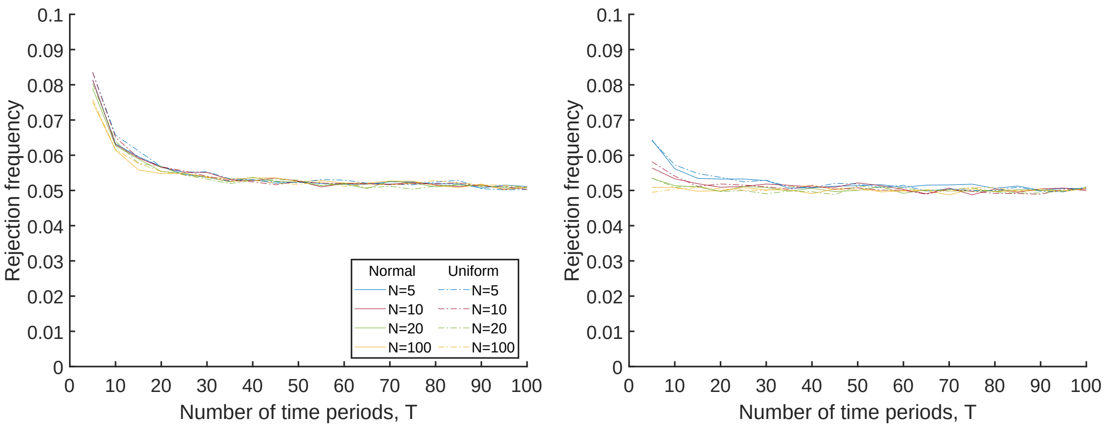
[20] If DGP parameter $\sigma^2_\alpha > 0$, we further know that $\text{Var}(\alpha_i) > 0$, $\rho_{\alpha v} \in (-1, 1)$, and that the correlation matrix of $[\alpha_i, v_i, \varepsilon^x_i]'$ becomes positive definite with determinant $1 + \rho^2_{\alpha x}\rho^2_{vx} - \rho^{d2}_{\alpha v} - \rho^2_{\alpha x} - \rho^2_{vx} > 0$.

DGP parameters and generate $N \times T$ observations.[21] We replicate this 100,000 times, redrawing all five innovations at each replication.

B.4.2 Size

Under the null hypothesis that all normalized country fixed effects α_i^0 in model (B.4) are zero, the non-normalized α_i do not depend on i, as shown in Section 3.1. The reverse is also true. Hence, to calculate the actual sizes of the Wald tests, we generate data under $\text{Var}(\alpha_i) = 0$.

Figure B.1: Actual sizes of Wald tests of the absence of fixed effects α_i^0



(a) Diagnostic test (b) Sensitivity test

Each gridpoint is based on 100,000 normal or uniform draws from DGP (B.5) using $\text{Var}(\alpha_i) = 0$, while the other base parameters turn out to be irrelevant. We estimate model (B.4) and compute the Wald tests of Appendices B.2 and B.3. Further details are in B.4.1.

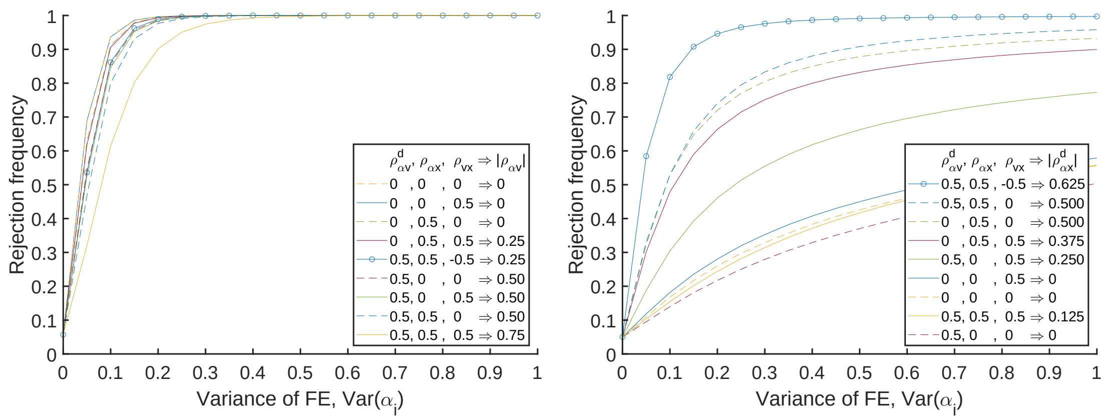
Figure B.1a presents the sizes of the Wald diagnostic test for various N and T. It is reassuring that the number of fixed effects, N, has virtually no effect on size. This could have been expected from Donald et al. (2003).[16] Size also hardly depends on the distribution of the innovations, as the solid line for the normal and the dashed line for the uniform distribution are so close. The test, however, is oversized for small T, say $T < 20$. This is in line with the cross-section results in Evans and Savin (1982). They suggest using the likelihood ratio test to avoid oversize, and our results for that test (not reported) corroborate that. Still, we keep our focus on Wald, because it is easier to use, oversize only makes our identification strategy conservative, and we have applications in mind with substantial T.

[21] If $\text{Var}(\alpha_i) > 0$, we get $\alpha_v = \rho_{\alpha v}^d \sqrt{\text{Var}(\alpha_i) / (1 - \rho_{vx}^2)}$, $\alpha_x = \rho_{\alpha x} \sqrt{\text{Var}(\alpha_i)}$, $\nu_x = \rho_{vx} / \sqrt{1 - \rho_{vx}^2}$, and $\sigma_\alpha^2 = \text{Var}(\alpha_i) \left[1 - \rho_{\alpha v}^{d2} / (1 - \rho_{vx}^2) - \rho_{\alpha x}^2\right]$, where the term in square brackets is positive because the determinant in footnote 20 is so. If $\text{Var}(\alpha_i) = 0$, the formula for ν_x remains, and $\alpha_v = \alpha_x = \sigma_\alpha^2 = 0$.

Figure B.1b shows the sizes for the sensitivity test. The distribution of the innovations has virtually no effect, and the test is better sized for small T than the diagnostic test.

B.4.3 Power

Figure B.2: Powers of Wald tests of the absence of fixed effects α_i^0 for $N = T = 20$



(a) Diagnostic test

(b) Sensitivity test

Each gridpoint is based on 100,000 normal draws from DGP (B.5) and base parameters below. We estimate model (B.4) and compute the Wald tests of Appendices B.2 and B.3. Further details are in B.4.1.
The five base parameters are as follows. First, under the alternative hypothesis, $\text{Var}(\alpha_i) > 0$, and we cover that by a grid from 0 to 1, using that beyond 1 power increases monotonically to unity. This grid is realistic, as in Section 5 we can compute the variance of the estimated FE to obtain some idea about realistic values of $\text{Var}(\alpha_i)$, and in a model without constant regressors we get 1.6, while a model with them gives 0.03.
Next, each of $\rho^d_{\alpha v}$, $\rho_{\alpha x}$, and ρ_{vx} has a grid $\{0, 0.5, -0.5\}$, giving 27 combinations. For nine of them the power curves are visible, because each of the other 18 curves coincides with an included curve.
Finally, $\sigma^2_\chi = 1$. The diagnostic test power is robust to changes in σ^2_χ. However, the power of the sensitivity test depends on it, not so much for $\sigma^2_\chi > 0.1$, but if σ^2_χ goes to zero, power drops a lot. In the latter case, x_{it} gets nearly time invariant, so in this sense comparable to v_i. Fortunately, in this case we still have the diagnostic test, which has high power. Note that for the two it-regressors in our empirical application in Section 5 we can estimate σ^2_χ-values, which gives 1.7 and 6.9, so our choice $\sigma^2_\chi = 1$ is sensible.

Based on the results of the previous section, we study the powers of both Wald tests for $N = 20$ and $T = 20$ with normally-distributed innovations.[22] Figure B.2 displays the powers as a function of the magnitude of the fixed effects, $\text{Var}(\alpha_i)$. Powers have not been size corrected. There are power curves for nine representative combina-

[22]This choice gives a representative view. Still, there are some moderate dependencies of power on panel sizes. The powers of both Wald tests depend positively on N, because for higher N the null hypothesis contains more constraints. The powers also depend positively on T, because higher T gives more observations to estimate each α_i^0.

tions $\left(\rho^d_{\alpha v}, \rho_{\alpha x}, \rho_{vx}\right)$, eight containing all combinations of ρs from $\{0, 0.5\}$, and one for $(0.5, 0.5, -0.5)$. The figure note motivates the choices of the base parameters.

Figure B.2a concerns the diagnostic test. For all ρ-combinations power quickly increases when the FE become stronger. Empirical Section 5 will present more evidence that the diagnostic test has serious power.

One can understand this power from the intuition in Ghysels et al. (2020). They argue that the power of Wald tests increases if regressors become less correlated. Our regressors are almost all dummies, which are orthogonal, and for orthogonal regressors the authors show that Wald works well. They derive this for hypotheses consisting of many constraints. In fact, in our setting, a larger N and thus more constraints makes Wald somewhat more powerful. Finally, their judgment is based on a Wald test that has lost power by adjusting for severe size distortion. We largely avoid such size distortion by using the degrees-of-freedom-corrected Wald test, as advised by Evans and Savin (1982), and that further explains the substantial power we find.

The ordering of the power curves yields insights into the power determinants, as follows. The top two items in the legend concern $\rho^d_{\alpha v} = \rho_{\alpha x} = 0$. The only source of the FE α_i is pure randomness ε^α_i, uncorrelated with observables. This is where the test has maximum power. Intuitively, the α_i that are hidden in y_{it} are neither explained by v_i, nor by x_{it}, so they are fully picked up by large (absolute) estimated α^0_i, causing high power. As this holds for any ρ_{vx}, the two top power curves coincide.

Going down the legend reduces power (slightly). The most important power indicator is the total correlation $\rho_{\alpha v} = \rho^d_{\alpha v} + \rho_{\alpha x}\rho_{vx}$ in the right column of the legend, which is monotonically linked to power. Another driver is $\rho_{\alpha x}$. Hence, correlations between the FE and both regressors matter, where stronger correlation lowers power.

Figure B.2b gives the power of the sensitivity test for the same nine ρ-combinations as before. This test reflects the influence of leaving out the FE on the estimated β, similar to the idea of an omitted variable bias test. It thus provides an indirect signal compared to the direct signal in the diagnostic test, so that it is not surprising that the sensitivity test has lower power. As before, power is increasing in $\text{Var}\left(\alpha_i\right)$, reflecting that the diagnostic part is relevant in the sensitivity test formula (B.3).

Because of the link to omitted variable bias, let us consider $\rho_{\alpha x}$. As usual, a high value by itself causes an upward bias in the estimated β, giving power to the test. From Basu (2020) we conclude that the indirect correlation term $\rho_{\alpha v}\rho_{vx}$ mitigates the bias. Hence, as before, we define the direct part of the correlation as $\rho^d_{\alpha x} = \rho_{\alpha x} - \rho_{\alpha v}\rho_{vx}$, and use this as a simple bias indicator. Its values are on the right side in the legend.

The legend is again ordered from high to low power, and the ordering confirms that the bias indicator is positively related to power, though not perfectly. We see that the lowest power occurs when there is low omitted variable bias, and that the test has substantial power when omitted variable bias is high. Both are reassuring.

C Robustness analysis

This appendix confirms that our results are robust to various deviations from the baseline specification. We focus on the identification tests, for which the precise normalization is irrelevant, but we also report the estimates under untangling normalization.

Table C.1: Sensitivity of results for exp_{iUSt} model (15)

Specification	5	7	8	9	10	11
Estimation	LSDV	No $\tau_i \cdot t$	DOLS	GPML	PPML	Pre-1979
gdp_{it}	1.12 *	2.26 *	0.46	1.12 *	0.25	1.39 *
	(0.25)	(0.19)	(0.33)	(0.25)	(0.36)	(0.20)
$reer_{it}$	0.66 *	0.83 *	0.50 *	0.67 *	0.36 *	0.28 *
	(0.15)	(0.21)	(0.16)	(0.15)	(0.10)	(0.12)
gdp_{USt}	2.35 *	1.55 *	2.81 *	2.26 *	3.05 *	2.27 *
	(0.39)	(0.50)	(0.33)	(0.39)	(0.38)	(0.40)
gdp_{Wt}	−1.18	−1.06	−1.21	−1.19	−2.46 *	−1.67 *
	(0.73)	(1.13)	(0.75)	(0.72)	(0.85)	(0.47)
$reer_{USt}$	−1.02 *	−0.97 *	−1.07 *	−1.01 *	−0.43 *	−1.26 *
	(0.13)	(0.17)	(0.13)	(0.13)	(0.18)	(0.10)
Wald tests						
$\theta_t^u = 0$	22.77	20.31	22.91	22.33	55.78 *	82.73 *
	[0.30]	[0.44]	[0.29]	[0.33]	[0.00]	[0.00]
$\beta = \beta\|_{\theta_t^u=0}$	0.67	3.24	0.94	0.55	2.95	0.79
	[0.72]	[0.20]	[0.62]	[0.76]	[0.23]	[0.68]
R^2_θ	0.98	0.97	0.98	0.98	0.97	0.95

All models include θ_t and 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 (B.3), 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. Model 11 is the same as 5, but the results are based on the enlarged 1965-2011 sample (765 observations), leading to 34 instead of 20 degrees of freedom for the Wald diagnostic test, giving critical value 48.60 instead of 31.41. The note to Table 1 provides further details.

C.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. More specifically, a Wald test of $\tau_i = 0$ for all i (leaving τ unrestricted) is 267, 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 C.1 to the baseline, replicated as 5. 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.

C.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 C.1 displays the results. They do not differ much from the baseline results in model 5. 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.

C.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 $\exp(\varepsilon_{it})$ and then use GPML or PPML. Models 9 and 10 of Table C.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 $\theta^u_t = 0$. However, the plot of estimated θ^u_t is similar to that of GPML (both not shown) and LSDV in Figure 2c, and only two are outside the band. We are thus still close to the LSDV conclusion, which is confirmed by the again high $R^2_\theta = 97\%$. The Wald sensitivity test does not reject, in line with LSDV.

C.4 Enlarged sample: 1965-2011

So far, we have considered a sample from 1979-2011. That has been sufficient to illustrate our contributions. Now, and only in this section, we add pre-1979 data, thereby including some economically unstable years.

Table C.1 specification 11 displays the results. The main difference with the baseline sample is that the Wald diagnostic test now rejects that all untangled time FE $\theta_t^u = 0$. However, the plot of estimated θ_t^u is similar to that of Figure 2c and only two are outside the band, reflecting that the time FE are almost completely explained by the t-regressors ($R_\theta^2 = 95\%$). Even the big economic swings before 1979 are captured quite well by the t-regressors. 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 $\omega^u = \omega$.

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 $\theta_t^u = 0$. Hence, 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.