Untangling Fixed Effects and Constant Regressors

Franc Klaassen
Rutger Teulings

Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute, the Netherlands.

Untangling Fixed Effects
and Constant Regressors

Franc Klaassen

University of Amsterdam and Tinbergen Institute

and

Rutger Teulings *

University of Amsterdam and Tinbergen Institute

December 23, 2015

Abstract

Fixed effects (FE) in panel data models overlap each other and prohibit the identification of the impact of "constant" regressors. Think of regressors that are constant across countries in a country-time panel with time FE. The traditional approach is to drop some FE and constant regressors by normalizing their impact to zero. We introduce "untangling normalization", meaning that we orthogonalize the FE and, if present, the constant regressors. The untangled FE are much easier to interpret. Moreover, the impact of constant regressors can now be estimated, and the untangled FE indicate to what extent the estimates reflect the true value. Our untangled estimates are a linear transformation of the traditional, zero-normalized estimates; no new estimation is needed. We apply the approach to a gravity model for OECD countries' exports to the US. The constant regressors US GDP, world GDP and the US effective exchange rate explain 90% of the time FE, making the latter redundant, so the estimated impacts indeed reflect the true value.

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

1 Introduction

In a panel data model one uses fixed effects (FE) to capture unobserved heterogeneity. In these models we have to take care of potential multicollinearity between two sets of dummy variables corresponding to the FE or between those dummy variables and the constant regressors, for example country FE and regressors that are constant over time. A common method to prevent multicollinearity is to set some fixed effects and impacts of constant regressors to zero; we call this zero normalization. All remaining parameters can now be estimated. But estimation of the impact of constant regressors is beyond reach.

We propose a new normalization method called untangling normalization. This method uses orthogonality between two types of FE and between FE and constant regressors as normalizations. These are normalizations, not restrictions as used in random-effects types of approach, so we stay fully within the FE paradigm. Computation of the untangled estimates is simply a linear transformation of the zero-normalized estimates without performing a second estimation. Untangling normalization has two improvements with respect to zero normalization, one regarding interpretation, and the other one concerning the estimation of the impact of constant regressors, which we now illustrate.

First, untangling normalization improves the interpretation by disentangling the combined explanatory power of the FE into separate sources, such as an overall mean and country deviations from that overall mean. Instead, zero normalization causes different sources to be 'tangled' in one FE. Our method thus simplifies interpretation of the FE.

For clarification, consider a country-time panel data model with a constant and country FE. The vector of ones is multicollinear with the country dummies, so we need one normalization. Zero normalization normalizes either the constant or any of the other country FE, say $i = 1$, to zero. In the first case, all country FE contain the overall constant. In the second case, the constant contains the overall mean and the country 1 effect, and all other FE are country deviations from country 1. The FE are 'tangled'. Untangling normalization here sets the mean of the country FE to zero, that is, it demeans the country dummies making them orthogonal to the vector of ones. This ensures that the constant captures the common mean and the country FE capture the country deviations from the mean; the effects are 'untangled' and assigned to a separate FE, easing interpretation[1]. Now, plotting them can help to obtain ideas of potentially

---

[1]This specific example is not new. Suits (1984) already sets the mean of individual effects to zero and uses it as a normalization to prevent multicollinearity. However, we consider a much more extensive

important regressors. So untangling normalization gives a unique normalization rule and improves interpretation.

The second improvement of untangling normalization is that it allows for estimation of the impact of constant regressors. The untangled fixed effects then contain what is left over. If that is small enough (as in our application), those fixed effects can safely be left out, so that the true impacts of constant regressors are identified. Such identification is not possible under zero normalization.

We clarify this with an example using a regressor that is constant over time. This constant regressor is multicollinear with the country dummies. Zero normalization usually normalizes the constant regressor parameter to zero and therefore does not exploit its explanatory power. However, untangling normalization orthogonalizes the country FE with respect to the constant regressor and the remaining country effects are captured by the country FE. This yields an estimate for the parameter of the constant regressor. If the country FE are now redundant, one can leave them out and the constant regressor captures all country-specific variation. This can lead to substantial efficiency gains as the number of constant regressors is typically much smaller than the number of countries.

The orthogonality condition we use is not an identifying restriction like in OLS, but just a normalizaiton. The price we pay for this is that the parameter of the constant regressor is not the true value but the pseudo-true value. However, the closer the untangled country FE are to zero, the closer the pseudo-true values are to the true values. Plotting the untangled FE with their confidence band gives a quick idea of how small the difference is. Because the estimation of the impact of constant regressors is a notorious problem in the literature, the untangled estimates are a worthwhile contribution.

The literature presents alternative methods, such as random effects (RE), fixed effects vector decomposition (FEVD) by Plümper & Troeger (2007), or the method by Hausman & Taylor (1981). These use orthogonality conditions as restrictions (not as normalizations like we do) to identify the parameters of constant regressors. To put it differently, all existing methods somewhere impose a random-effects type of restriction. In contrast, our approach does not a priori assume identification of the true values and is a fully fixed effect approach. For each method we will now elaborate on this.

The RE method imposes additional restrictive moment conditions between the (stochastic) effects and the it-regressors, regressors varying both over time and countries. This is not needed for untangling normalization and is typically the reason why

---

FE-configuration and we even use it to untangle constant regressor parameters and FE.

FE is preferred over RE.

The FEVD is a two-step estimation method; the first step applies FE using zero normalization, and the second step regresses the estimated fixed effects on the constant regressor, and the error term is assumed uncorrelated with the constant regressor, so the error term is like a RE. It uses weaker orthogonality conditions than RE, because it does not need orthogonality conditions between the effects and the it-regressors. The two-step nature of FEVD has three consequences. First, FEVD does not give standard errors for the estimated effects (the residuals in the second step) because they are considered to be random. Untangling normalization gives the standard errors of each country FE. These are important to evaluate the difference between the pseudo-true and the true value of the constant regressor impact. Second, the estimated covariance matrix in the second step needs to be adjusted (e.g. Greene (2011)). Untangling normalization gives the correct covariance matrix straight away, because it is just a linear transformation of the zero-normalized covariance matrix. This is simpler. Finally, in FEVD it is not obvious what to include on the left- and right-hand sides in the second-step regression, as Pesaran & Zhou (2014) emphasize.

Hausman & Taylor (1981) propose an instrumental variables method to estimate the impact of constant regressors. They propose three different types of instruments: variables from outside the model, it-regressors, or constant regressors from the model. All types of instruments are assumed to be uncorrelated with the common effects of interest, which are treated as random variables. For example, in case country-specific regressors are used as instrument, they are uncorrelated with the country effect, which is the same restriction as in FEVD; see Breusch et al. (2011a) for a detailed analysis. Our method does not rely on external instruments or such RE type of restrictions.

We apply untangling normalization in a gravity model for exports from 17 OECD countries to the US. We show that the untangled time FE are indeed easier to interpret. They reveal the time-specific development that is missing from the included it-regressor and suggest the importance of three constant regressors, namely US gdp, world gdp and US effective exchange rate. The gravity literature suggests that these three constant regressors matter as well, but the problem is that their effects cannot be estimated using the zero-normalized FE approach. Untangling normalization allows us to exploit these three constant regressors and estimate their parameters, albeit that they are the pseudo-true values. The three constant regressors explain 90% of the time deviations. Investigating the remaining untangled time FE demonstrates that their contribution to the model is redundant. This implies that the estimated constant regressor parameters indeed match the true values, a result that could not have been established with zero

4

normalization.

The paper is organized as follows. We start by introducing the model and discuss the identification problem in Section 2. In Section 3 we discuss zero normalization and how we can estimate the identified model. Section 4 introduces our new approach, untangling normalization. In Section 5 we show how to transform zero-normalized estimation results to any general normalization that is linear in the parameters, and apply it to untangling normalization. We apply untangling normalization in Section 6 to a gravity model. Finally, Section 7 concludes.

2   Model specification and identification problem

In this section we introduce the model specification that we use throughout the paper. We consider a two-dimensional balanced panel data model with dimensions i (e.g. country) and t (e.g. time); generalization are left for future research. There are N countries and T time periods. All vectors are column vectors unless stated otherwise.

The dependent variable y_{it} is modeled by

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

where ε_{it} is the error term. This model contains two main features. The first one is the inclusion of deterministic variables, the different fixed effects (FE) types. We have three FE-families, each targeting a specific dimension of the data. The α-family targets the mean a cross countries. It has a homogeneous type, denoted by the parameter α, and a heterogeneous type, denoted by α_i. The τ-family targets the linear trends across countries. It has a homogeneous type, τ, and a heterogeneous type τ_i. Finally, the θ-family targets the variation over time. It only has a heterogeneous type, θ_t. In total we have K_d FE.

The second feature of model (1) is that it has i-, t-, and it-regressors. The vector v_i of length K_v depicts all variables that are constant over time. The vector w_t of length K_w captures all variables constant over countries. Finally, all variables that vary over countries and time are represented in the vector x_{it} of length K_x.

It is sometimes convenient to use (1) in matrix-form, where the time series of the countries are stacked. All K_d deterministic variables are contained in the deterministic D matrix and the FE parameters in vector $\delta = [\alpha, \alpha_1, \ldots, \alpha_N, \tau, \tau_1, \ldots, \tau_N, \theta_1, \ldots, \theta_T]'$. Both country and time-specific variables are stacked into Z, such that row it becomes $[v_i', w_t']$, consisting of in total $K_z = K_v + K_w$ regressors and its corresponding parameter vector is $\gamma = [\nu', \omega']'$. Finally, all observations of variables varying both over countries

and time are stacked in X. We assume that all regressors in Z and X are linearly independent from all other regressors in Z and X, that is, $[Z, X]$ has full column rank. Hence, the model in matrix notation becomes

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

where y and ε stack all y_{it} and ε_{it}, respectively.

Not all parameters are identified due to multicollinearity. For example, increasing α by a number and decreasing all α_i by the same number yields the same value of y. So the first $1 + N$ columns in D are linearly dependent. Also the presence of both v_i and country dummies and both w_t and time dummies lead to multicollinearity. Therefore, we have to introduce several normalizations to ensure all parameters are identified, and only then we can estimate the model.

3   Traditional approach: zero normalization and estimation

3.1   Zero normalization

The easiest and most common way to circumvent multicollinearity in (2) is by zero normalization, which sets parameters to zero. The multicollinearity problem only manifests itself in $[D, Z]$, because all variables in X, varying over countries and time, are linearly independent from D, Z and each other. We first discuss zero normalization due to multicollinearity in D and later include Z. To denote zero-normalized parameters we put a 0 superscript.

Consider, for example, a panel data model with α and α_i. To circumvent multicollinearity of the variables underlying α and α_i, we need one normalization. Most software packages use zero normalization, but this does not offer us a unique normalization. Typical choices are to normalize α or one specific α_i to zero. If we normalize α to zero, the country FE parameter α_i^0 is the overall mean plus some country effect. If we normalize α_i to zero for $i = 1$, say, α^0 is the overall mean plus the effect of country 1, and all remaining α_i^0 represent the country deviations from country 1. In either way one column in the regressor matrix can be dropped, so that zero normalization ensures that the stripped regressor matrix has full column rank such that the rest of the parameters of the model can be estimated.

This example reveals the first disadvantage of zero normalization. The interpretation of the parameters differs over the different zero normalizations and in all cases

interpretation is difficult because different effects are 'tangled' in the remaining parameters.

In general, the dummy matrix D has column rank $K_d - m_d$, where m_d is the number of multicollinearities, i.e. dependent columns, within D. The value of m_d depends on the FE-configuration. To prevent multicollinearity in D we need to perform m_d zero normalizations in δ. The K_d vector containing the m_d zero-normalized FE parameters and all other $K_d - m_d$ FE parameters is denoted by δ^0. We distinguish δ^0 from δ to highlight that it has been identified using the zero normalization.

There is not only potential multicollinearity within the D matrix, but also between D and Z. If the model contains country FE α_i and country-specific regressors v_i, then the country dummies span v_i, leading to multicollinearity. The common approach is to normalize K_v parameters to zero to ensure that after stripping the regressor matrix has full column rank. The most common way is to normalize all K_v elements in ν to zero. The corresponding zero-normalized parameter vector is denoted by $\nu^0 = 0$.

Similarly, time dummies and w_t variables are multicollinear. If both are present in the model, one typically normalizes all K_w elements in ω to zero and the corresponding zero-normalized parameter vector is denoted by $\omega^0 = 0$. If ν or ω is normalized to zero, γ^0 is the encompassing zero-normalized parameter vector. The number of multicollinearities between Z and D is given by m_z and depends both on the FE- and Z-configuration in the model.

This example shows the second disadvantage of zero normalization; one does not exploit the explanatory power of the constant regressors.

In summary, the column rank of the complete regressor matrix in (2), $[D, Z, X]$, is $K_d + K_z + K_x - m_d - m_z$ and $m_d + m_z$ parameters in $[\delta', \gamma']'$ are normalized to zero. We emphasize two properties of zero normalization, namely the interpretation differs over different zero normalizations and is cumbersome in all cases, and it cannot exploit the explanatory power of the constant regressors. The attractiveness of zero normalization is that parameters are easy to estimate. We now discuss estimation that exploits this, after which we will show how our approach improves regarding the two disadvantages.

3.2 Estimation

If we normalize a parameter to zero, we in practice just omit the corresponding variable in the regressor matrix. This results in a compact set of $K_d + K_z + K_x - m_d - m_z$ independent regressors. The corresponding stripped parameter vector has the same length.

It does not matter with what kind of estimation method we estimate the parameters,

as long as the FE and constant regressors enter linearly into the explanatory part of the model. For example, if the regressors are uncorrelated with the error term, one typically uses the least squares dummy variable (LSDV) estimator. But in the gravity model application, we will show that our approach also works if one prefers imposing other assumptions and using Poisson pseudo maximum likelihood for estimation, as proposed by Santos-Silva & Tenreyro (2006).

After estimation, we place back the zero-normalized parameters such that we obtain $\widehat{\delta}^0$ and $\widehat{\gamma}^0$ of length K_d and K_z, respectively. We insert rows and columns of zeros into the estimated covariance matrix corresponding to the zero-normalized parameters such that we get $\widehat{V}^0$. We now have zero-normalized estimates for all parameters and the corresponding covariance matrix.

4 New approach: untangling normalization

4.1 General idea

The general idea of untangling normalization is to prevent multicollinearity between two sets of variables by making the parameters orthogonal to each other and use this as normalization. We can now interpret the parameters of the first set as a deviation on the second set of parameters; we have untangled the parameter estimates. We use two examples to get a better understanding of our approach and illustrate the improvements of untangling normalization with respect to zero normalization. Untangled parameters are denoted with a u superscript.

Our first example focusses on multicollinearity within the deterministic D in model (2), so we ignore Z for now. We use the simple panel data model of Section 3.1, a model with a constant α and country FE α_i. We need one normalization.

Zero-normalization causes the overall mean and the country effects to be 'tangled' into the estimated FE, as we have shown in Section 3.1. Untangling normalization sets the mean of the country FE α_i to zero. Now, the untangled constant α^u, the homogeneous FE-type, captures the common mean, and the untangled country FE α_i^u, the heterogeneous FE-type, captures the country deviations from the mean. So both effects do not interfere with each other and are assigned to separate parameters, easing interpretation. This can help us to identify potentially important regressors based on the estimated untangled α_i^u.

Note that it is important to include a constant in the model, otherwise α_i would also capture the common mean, which would hamper interpretation.

In the second example we introduce multicollinearity between D and Z by adding

8

the variables v_i to our first example. This leads to multicollinearity between the country dummy variables and v_i, and to prevent this we need K_v normalizations in addition to the one above.

Zero normalization normalizes all K_v parameters in ν to zero, so the information in the constant regressors v_i is not used. To be able exploit the constant regressors v_i we use untangling normalizations. They set all variables in v_i orthogonal to the country FE, like the OLS moment condition that sets regressors orthogonal to the error term. In this way the untangled parameters α^u and ν^u are identified, and the α_i^u capture the remaining country effects orthogonal to v_i and the vector of ones[2].

In general, the untangled parameter ν^u is not the true but the pseudo-true value of the impact of v_i. However, if the remaining country effects α_i^u are zero, then leaving them out does not affect the error term, so it will not cause omitted variable bias, and ν^u becomes the true value. Hence, if we capture all country effects by adding v_i, we estimate the true value of ν. Because we have estimates of α_i^u and know their estimated variance, we can test whether it is safe to leave out the country FE and interpret the estimated ν_i^u as true values. Dropping country FE can lead to substantial efficiency gains if the number of countries is larger than the number of constant regressors.

In the subsequent part of this section we will discuss untangling normalization for each FE-type and constant regressor separately. Finally, we will discuss similarities of untangling normalization with alternative estimation methods.

4.2 Untangling fixed effects

We start by introducing all untangling normalizations for different FE-types.

Constant: α

Because the overall constant α is the most homogeneous FE-type in the FE configuration, we do not normalize it. We want it to capture the overall mean of all observations.

Country-specific effects: α_i

We need one normalization to prevent multicollinearity between country dummies and the vector of ones. To untangle the country FE α_i from the constant α, we normalize

---

[2]It is also possible to add Mundlak (1978) terms to v_i, that is, the mean of x_{it} over time, denoted by $\bar{x}_{i.}$. This allows for an even more general model specification. In our application in Section 6, the addition of Mundlak terms does not influence our results because the mean over countries of all x_{it} variables is already proxied by an actual economic variable in w_t. For simplicity we ignore Mundlak terms here.

the mean of α_i to zero, so that the untangled α_i^u capture the country deviation from the overall mean α^u

$$\sum_i \alpha_i^u = 0. \tag{3}$$

Common trend: $\tau \cdot t$

Just as α, we do not normalize the common trend parameter τ, irrespective of any FE-configuration. We want it to capture the common trend of all observations and there is no more homogeneous trend FE-type than $\tau \cdot t$.

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

Similar to α_i, we perform one normalization on the country trend variable because there is a common trend variable in the model. We normalize the mean of the country trend FE $\tau_i \cdot t$ to zero, so that the untangled $\tau_i^u \cdot t$ captures the country deviation upon the common trend $\tau^u \cdot t$

$$\sum_i \tau_i^u = 0. \tag{4}$$

Time-specific effects: θ_t

Finally, we have the time FE θ_t. We need one normalization to prevent multicollinearity between the time dummies and the vector of ones. We normalize the mean of θ_t to zero, such that the untangled time effects θ_t^u are untangled from α^u and α_i^u, and they can be interpret as the time deviation from the overall mean:

$$\sum_t \theta_t^u = 0. \tag{5}$$

Time FE θ_t also pick up trends. Because we have a common trend in our model, this leads to multicollinearity between time dummies and the trend variable. To prevent multicollinearity and untangle θ_t from $\tau \cdot t$ and $\tau_i \cdot t$, we orthogonalize θ_t with respect to the trend. This ensures that θ_t^u is trendless and can be interpreted as time deviation from the common trend. Hence, the second untangling normalization for θ_t becomes

$$\sum_t \theta_t^u \cdot t = 0. \tag{6}$$

In summary, by choosing normalizations smartly we manage to untangle all the different FE-types from each other. In this way each FE-type has a separate function in the model, easing interpretation.

4.3 Exploiting constant regressors

We can also apply our untangling normalization method to estimate both the impact of non-deterministic variables (e.g. v_i) and a FE-type (e.g. α_i), both varying solely over the same dimension.

Country-specific regressors: v_i

To prevent multicollinearity between v_i and the country dummy variables and untangle ν from α_i we need K_v untangling normalizations. We project the country effects on the null space of v_i and use these orthogonality conditions as untangling normalizations. Hence, we get the following K_v normalization conditions, one for each regressor k in v_i:

$$\sum_i \alpha_i^u v_i^k = 0. \tag{7}$$

Time-specific regressors: w_t

In a similar way, we have to prevent multicollinearity between w_t and the time dummy variables. We normalize and untangle the parameters ω from θ_t by projecting the time dummies on w_t resulting into K_w normalizations, one for each regressor k in w_t:

$$\sum_t \theta_t^u w_t^k = 0. \tag{8}$$

4.4 Differences with other methods.

Existing approaches like random effects (RE) and fixed effects vector decomposition (FEVD) use similar orthogonality conditions as our approach. The key difference is that in our method the orthogonality conditions are just normalizations and not restrictions. In other words, all these other methods impose at some point types of RE restrictions, while untangling normalization just applies FE. In the remainder of this section we will compare our untangling normalization method with respect to existing methods in more detail.

Zero-normalized fixed effects

Untangling normalization is based on FE. Hence, the estimates $\widehat{\beta}$ for the it-regressors are identical whatever the normalization, the traditional zero or our untangling normalization. However, untangling normalization allows us to better interpret the FE estimates. This allows us to identify potentially important constant regressors and exploit their explanatory power, as explained in Section 4.

Random effects

In a RE approach, the effects, such as α_i, are random variables. To estimate (2) with RE, one typically imposes the following two moment conditions: $\mathbb{E}\{x_{it}\alpha_i\} = 0$ and $\mathbb{E}\{v_i\alpha_i\} = 0$; similar moment conditions are needed for the other effect types. This is where untangling normalization differs from RE. Our method does not need orthogonality of α_i and x_{it}. Moreover, we do not impose orthogonality of v_i as a restriction, but only as a normalization. The FE nature of our approach makes it more robust, which is why FE is often preferred over RE. Still, if all RE restrictions happen to hold, the RE estimator will be more efficient.

Fixed effects vector decomposition

The fixed effects vector decomposition (FEVD) by Plümper & Troeger (2007) is a two-step estimation method. First, it estimates the FE parameters using the same methods as described in Section 3. The second step uses the vector of estimated FE (e.g. α_i) as dependent variable and explains it with constant variables (e.g. v_i), treating the remainder as a random variable, that is, a RE. The residuals are estimates of the effects left unexplained by the constant regressor.

This two-step method differs from untangling normalization. First, unlike our approach, FEVD does not give standard errors for each remaining country effect separately, because they are the residuals of the second estimation. These standard errors are important to evaluate the difference between the true and pseudo-true values of the impacts of the constant regressors.

Second, the standard errors obtained in the second regression need to be adjusted, as argued by Breusch *et al.* (2011b) and Greene (2011), who also propose similar corrections. Untangling normalization does not perform a second regression. Instead, the untangled parameter estimates and standard errors just follow from a simple linear transformation of the estimated zero-normalized estimation results. So the untangled normalized standard errors need no correction as with FEVD.

Third, it is not clear what to include on the right- and left-hand sides (RHS and LHS, respectively) in the second-stage regression. On the RHS of the second-stage regression Plümper & Troeger (2007) only include w_t. Pesaran & Zhou (2014) show that it is important to also include a constant to avoid estimation bias. Conditional on having a trend, this raises the question whether one also has to include the overall trend $\tau \cdot t$ on the RHS, a question that we will explore and answer positively in the application section.

On the LHS Plümper & Troeger (2007) include the estimated α_i, but one could

12

also consider including the estimated constant. In more extensive models like (1), the number of choices becomes even larger. All these different choices for both the RHS and LHS make it difficult for the empirical researcher, and it is easy to make mistakes. Untangling normalization avoids such choices, because it transforms all estimated fixed effects in a unique way.

5 Implementing a renormalization

We have discussed zero normalization, estimation, and untangling normalization. This section will show how to transform the parameters δ^0 and γ^0 into another general-normalized parameter set δ^g and γ^g (containing the untangled parameters δ^u and γ^u as a special case) as long as the normalization is linear in the parameters[3]. Renormalization is just a redistribution of the total effect $D\delta^0 + Z\gamma^0$ over $K_d + K_z$ general-normalized parameters, resulting in $D\delta^g + Z\gamma^g$; no new estimation needed. All we need is a set of equations to solve for the unknown parameter vector $[\delta^{g\prime}, \gamma^{g\prime}]'$. Note that β does not change due to the renormalization, because the columns in X are linearly independent from $[D, Z]$ and therefore we do not need to (re)normalize β. We will first show renormalization in general and next we will zoom in into untangling normalization.

5.1 General linear renormalization equations

To obtain the renormalized parameter vector $[\delta^{g\prime}, \gamma^{g\prime}]'$ from the zero-normalized parameter vector $\left[\delta^{0\prime}, \gamma^{0\prime}\right]'$, we need to have at least $K_d + K_z$ independent equations to solve the system of equations for $[\delta^{g\prime}, \gamma^{g\prime}]'$.

We obtain the first set of equations from the fact that after renormalization the total value explained for each observation must stay the same. Hence, we obtain

$$D\delta^g + Z\gamma^g = D\delta^0 + Z\gamma^0. \tag{9}$$

These are NT equalities, but only $K_d + K_z - m_d - m_z$ are actually independent equations because that is the column rank of the regressor matrix $[D, Z]$, as derived in Section 3.1. Hence, for the system of equations to be fully identified we need at least $m_d + m_z$ additional equations. For this we use the zero- and g-normalizations.

First, we introduce the zero normalization matrix N^0 that selects the $m_d + m_z$ zero-normalized parameters. For example, consider the simple model without Z and only a constant and country FE, depicted in D, so we need only one normalization. If

---

[3]It is also possible to start the transformation from any other set of general-normalized parameters, as long as these normalizations are linear in the parameters as well.

we normalize the constant to zero, we get $N^0 = [1, 0, 0, .., 0]$, where N^0 is a $1 + N$ row vector. Instead, if we normalize country FE i to zero, we get $N^0 = [0, .., 0, 1, 0, .., 0]$, where the $1 + i$th element in N^0 equals one. Likewise, in our general model, a row in N^0 has a 1 at the place corresponding to a zero-normalized parameter. The zero normalization in matrix-form thus becomes

$$N^0 \begin{bmatrix} \delta^0 \\ \gamma^0 \end{bmatrix} = 0. \tag{10}$$

Next, we setup the g-normalization matrix N^g containing all $m_d + m_z$ normalizations belonging to the g-normalization. It is[4]

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

Combining (10) and (11) gives $m_d + m_z$ extra equalities. Together with (9) the number of independent equations is now equal to the number of unknown parameters in $[\delta^{g\prime}, \gamma^{g\prime}]'$ and we can solve for the latter. The system of equations can be written as

$$\begin{bmatrix} D & Z \\ & N^g \end{bmatrix} \begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = \begin{bmatrix} D & Z \\ & N^0 \end{bmatrix} \begin{bmatrix} \delta^0 \\ \gamma^0 \end{bmatrix}. \tag{12}$$

5.2 Solving for general-normalized parameters

We simplify (12) to

$$R^g \begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = R^0 \begin{bmatrix} \delta^0 \\ \gamma^0 \end{bmatrix}, \tag{13}$$

where the definitions of R^g and R^0 follow from (12). To solve for $[\delta^{g\prime}, \gamma^{g\prime}]'$, we start by pre-multiplying (13) with $R^{g\prime}$. The resulting matrix on the left-hand side of the expression becomes a square matrix with full column rank $K_d + K_z$ and is thus invertible. Hence,

$$\begin{bmatrix} \delta^g \\ \gamma^g \end{bmatrix} = R^{0g} \begin{bmatrix} \delta^0 \\ \gamma^0 \end{bmatrix}, \tag{14}$$

---

[4]We can generalize (11) by allowing the right-hand-side to be a non-zero vector c^g, where all elements can be any number. All subsequent expression can be easily changed accordingly. However, for ease of notation we restrict the right-hand-side of (11) to be equal to 0.

where R^{0g} is the renormalization-matrix that converts zero into general normalization[5]

$$R^{0g} = \left(R^{g\prime}R^g\right)^{-1} R^{g\prime}R^0. \tag{15}$$

It only consists of observables.

To obtain the g-normalized covariance matrix $V_{\delta\gamma}^g$ corresponding to δ^g and γ^g we use R^{0g} to transform the zero-normalized covariance matrix $V_{\delta\gamma}^0$ regarding δ^0 and γ^{0}[6]

$$V_{\delta\gamma}^g = R^{0g}V_{\delta\gamma}^0 R^{0g\prime}. \tag{16}$$

An advantage from a practical point of view is that no additional estimation or standard error correction is needed.

5.3 Renormalization applied to untangling normalization

To solve for the untangled normalized parameters, we can just use (14) except for replacing N^g in R^{0g} by the N^u matrix. We show how to construct the N^u matrix based on the untangling normalizations in Section 4. The order of parameters in N^u is the same as in (1). Hence, the untangling normalization matrix N^u consists of seven blocks corresponding to the parameters involved in the normalization $[\alpha, \alpha_i, \tau, \tau_i, \theta_t, \nu', \omega']$, where each block has the following number of columns $[1 \; N \; 1 \; N \; T \; K_v \; K_w]$. Subsequently, we suppress the length of each block for ease of notation and we use 0 and 1 to denote blocks of zeros and ones, respectively. We assume that all FE-types and variables v_i and w_t are present in the model. If not, its corresponding block must be omitted from the N^u matrix.

For example, the first row of N^u, where we normalize the mean of α_i to zero because of the constant, is $[0, 1, 0, 0, 0, 0, 0]$, where the 1 is a row of ones with length N, as explained above. Indeed, if we combine this with (11), we get (3).

---

[5]The renormalization-matrix R^{0g} depends on the matrix multiplication $R^{g\prime}R^g$. Multiplying R^g with itself can lead to large or small numbers, and for numerical reasons it might be better to avoid this. A solution is to use Gaussian elimination to select all $K_d + K_z$ independent rows in R_g, making the resulting $\widetilde{R}_g$ a square matrix of full rank. To maintain the equalities, we select the same rows in R_0 and obtain the square matrix $\widetilde{R}^0$. We can replace R^g and R^0 in (15) by $\widetilde{R}^g$ and $\widetilde{R}^0$, respectively, and get the $\widetilde{R}^{0g} = \left(\widetilde{R}^g\right)^{-1} \widetilde{R}^0$, which we then use in (14) instead of R^{0g}. So we have avoided the multiplication of R^g with itself and this method yields the same g-normalized parameters as when we use (15) as renormalization matrix.

[6]In some cases it might be important to also obtain the covariances corresponding to β and δ^g or γ^g. Then we need to extend (13) by appending $[X', 0']'$ to the right of both R^0 and R^g and extending the g- and zero-normalized parameter vectors with β, such that the renormalization-matrix R^{0g} has dimension $K_d + K_z + K_x$. Now we can transform the complete zero-normalized covariance matrix V^0 to V^g, similar to (16), obtaining the complete g-normalized covariance matrix.

We get the following untangling normalization matrix based on normalizations (3)-(8):

$$
\begin{array}{cccccccc}
 & \alpha & \alpha_i & \tau & \tau_i & \theta_t & \nu' & \omega' & N^u \text{ row implements}
\end{array}
$$

$$
N^u = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1...T & 0 & 0 \\
0 & v_1...v_N & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & w_1...w_T & 0 & 0
\end{bmatrix} ;
\quad
\begin{array}{l}
\sum_i \alpha_i^u = 0 \\
\sum_i \tau_i^u = 0 \\
\sum_t \theta_t^u = 0 \\
\sum_t \theta_t^u \cdot t = 0 \\
\sum_i \alpha_i^u v_i^k = 0 \\
\sum_t \theta_t^u w_t^k = 0
\end{array}
. \qquad (17)
$$

We see that the complete two blocks corresponding to γ consists of zeros. So there are no normalizations on γ.

6 Application: untangling the gravity model

We can apply the new untangling normalizaton method to various economic models to obtain parameter estimates which used to be infeasible before due to the added FE. Take for example the gravity model.

6.1 The gravity model

The general idea of the gravity model is that export between two countries depends positively on both the exporting and the importing country's GDP divided by world GDP and negatively on distance, where distance can be both physical and economic distance between countries. One proxy for economic distance is the trade costs between two countries.

The most common gravity model is that of Anderson & Wincoop (2003). They show that, besides bilateral trade costs, it is important to include multilateral resistance terms for the importer and exporter to avoid estimation bias. However, the problem is to find economic variables that can serve as a proxy for these terms.

One approach to nevertheless control for the multilateral boundaries, is to introduce country-time FE. A disadvantage is that the underlying dummies are perfectly multicollinear with country- and time-specific variables, such as exporter's, importer's and world GDP, and the corresponding parameters are therefore not identified.

As an alternative, we apply our untangling normalization method. Afterwards we can test whether the remaining FE still matter. If not, they can be left out and we have estimated the previously unidentified parameters.

6.2   Model specification

Typical panel data gravity models consider exports from country i to j in year t. To keep our application simple and in line with previous sections, we only look at export towards one country j, the US, reducing the dimensions of the data from three to two (see Klaassen & Teulings (2015) for a three dimensional application). We add exporter, US and world GDP, as suggested by the gravity model. Furthermore, we use the theory from Klaassen & Teulings (2015) as a motivation to add the bilateral exchange rate and also the exporter and US effective exchange rates, where the latter two serve as proxy for the exporter's and US' multilateral boundary, respectively. All variables are in logs and real terms.

We add the general set of FE. Given our two-dimensional approach, these are the exporter and time FE and the exporter trend FE. The exporter effects control for distance between the exporting country and the US, common language with the US, and so on. Time effects capture global and US-specific developments, such as oil shocks, the recent financial crisis, and US dollar swings. Exporter trend effects control for the strong trends in exports and GDP, reflecting the typical approach in time series modeling, and following Bun & Klaassen (2007) and Baier *et al.* (2014).

Hence, the model becomes:

$$
\begin{aligned}
exp_{iUSt} =& \beta_1 gdp_{it} + \beta_2 rer_{iUSt} + \beta_3 reer_{it} + \gamma_1 gdp_{USt} + \gamma_2 gdp_{Wt} + \gamma_3 reer_{USt} \\
&+ \alpha + \alpha_i + \tau \cdot t + \tau_i \cdot t + \theta_t + \varepsilon_{it},
\end{aligned}
\tag{18}
$$

where exp_{iUSt} represents real export from country i to the US in year t, gdp_{it}, gdp_{USt} and gdp_{Wt} are real GDP of country i, the US and the world, respectively, rer_{iUSt} is the real exchange rate between country i and the US (the number of goods basket of country i to be paid for one US basket) and finally $reer_{it}$ and $reer_{USt}$ are the real effective exchange rates (reer) of country i and the US, respectively. So we have three it-regressors, which vary over both countries and time, and three t-regressors, which are constant across countries but vary over time. We assume that all explanatory variables are uncorrelated with the error term ε_{it}. We allow ε_{it} to be heteroscedastic and serially correlated and we assume it is not cross-sectionally correlated.

Specification (18) is sufficient to illustrate our method. Our main results are robust against different model specifications, such as adding dynamics, omitting country trend FE, and accounting for non-stationarity and cointegration, as we will show.

In the rest of this section we compare the estimated zero-normalized time FE θ_t^0 with the estimated untangled time FE θ_t^u. We use the estimated θ_t^u and gravity theory to come up with three t-regressors, US GDP, world GDP and US reer, that can explain

the estimated θ_t^u. We investigate how much of θ_t^u can be explained by these regressors and whether the remaining θ_t^u can safely be left out.

6.3 Data

We focus on $N = 17$ countries and their exports to the US, namely the EU-15 countries except for Belgium and Luxembourg[7], Canada, Japan, Norway and Switzerland. Our data run from 1965-2011 ($T = 47$) resulting into 799 observations.

We use monthly nominal export data from the IMF Direction of trade statistics (DOTS) and convert it from US dollars into home currency using the monthly nominal exchange rate from the International Financial Statistics (IFS) of the IMF. We average the converted nominal export to get yearly averages. To change these into real data we use the yearly export price index from the AMECO database provided by the European Commission.

Nominal yearly GDP data are from AMECO, where we us the exchange rate from AMECO to denominate it in the national currency. To construct real GDP data we use the AMECO GDP deflator. West-German GDP is used as a proxy for German GDP before 1991. We take real world GDP from the OECD Economic Outlook.

We construct the real exchange rate with the monthly nominal exchange rate from the IFS. First, we calculate yearly averages. Next, we multiply the nominal exchange rate by the ratio of the US GDP deflator over the GDP deflator of country i.

Finally, we use consumer-price-based monthly reer data from the Bank for International Settlements (BIS) and construct yearly averages. We invert the outcome such that an increase in reer coincides with a depreciation of the currency.

## 6.4 Zero normalization and estimation

Before we can estimate static model (18), we implement zero normalization to prevent multicollinearity. We normalize the last country FE α_N and trend FE τ_N to zero to break the dependence between α and α_i and between τ and τ_i. To prevent multicollinearity regarding the time FE and α and τ we normalize the latter two to zero. Finally, we normalize all γ to zero.

Unless reported otherwise, we use the LSDV estimation method. We apply Newey-West standard errors with optimal lag structure (see Newey & West (1987, 1994)) to correct for heteroscedasticity and serial correlation in ε_{it}. The estimation method is

---

[7]Austria, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden, United Kingdom

Figure 1: Two different zero normalizations and their impact on the time FE θ_t^0.
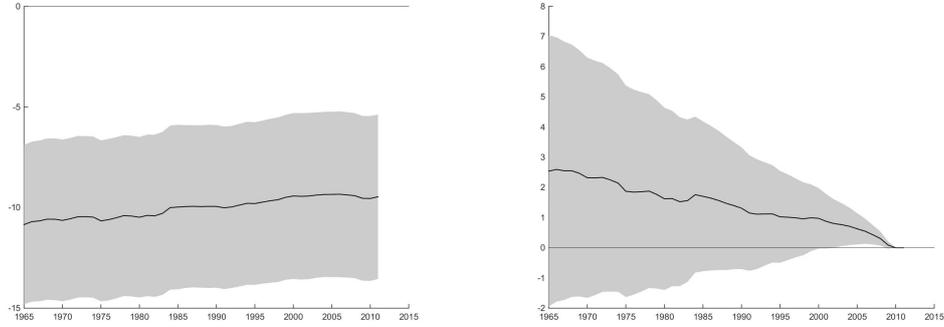
(a) $\alpha = \tau = 0$ (and $\alpha_N = \tau_N = 0$).

(b) $\theta_{T-1} = \theta_T = 0$ (and $\alpha_N = \tau_N = 0$).

Table 1: Estimation results for exp_{iUSt} based on (18).

Include θ_t	No			Yes		
DL-extension	No	No	Yes	No	No	Yes
Model	1	2	3	4	5	6
gdp_{it}	1.61*	1.15*	1.38*	1.23*	1.23*	1.40*
	(0.17)	(0.18)	(0.19)	(0.20)	(0.20)	(0.20)
rer_{iUSt}	1.05*	0.74*	0.76*	0.58*	0.58*	0.71*
	(0.07)	(0.15)	(0.20)	(0.21)	(0.21)	(0.22)
$reer_{it}$	-0.64*	-0.51*	-0.48*	-0.34	-0.34	-0.43
	(0.17)	(0.20)	(0.23)	(0.24)	(0.24)	(0.25)
gdp_{USt}		3.02*	2.29*		2.95*	2.25*
		(0.30)	(0.38)		(0.30)	(0.39)
gdp_{Wt}		-0.54	-1.38*		-0.62	-1.34*
		(0.40)	(0.47)		(0.39)	(0.47)
$reer_{USt}$		-0.32	-0.44		-0.50*	-0.54*
		(0.20)	(0.24)		(0.24)	(0.27)
F-test $\theta_t^u = 0$	–	–	–	7.85	1.78	1.26
p-value	–	–	–	0.00	0.00	0.15

Model 3 and 6 display the long-run effects estimates in a DL model including two lags of each regressor. Standard errors are between brackets. * indicates significance at 5% level.

more extensively described in Section 3. We use a 5% significance level throughout the paper.

Figure 1a displays the estimated zero-normalized time FE θ_t^0 and the confidence band. The mean is highly negative, reflecting that the θ_t^0 are disturbed by the overall means of the dependent and explanatory variables, as we have normalized α and α_N to zero. The θ_t^0 exhibit some variation over time but, this is hard to analyze.

Instead of the constant and common trend, we can also normalize two time FE to zero, for example the last two, as is displayed in Figure 1b. This leads to downward sloping and mostly insignificant θ_t^0 estimates with a positive mean. This complete change compared to Figure 1a exemplifies the impact different zero normalizations have on the FE estimates.

Table 1 presents the estimation results. The estimates of Model 4 underly the figures just discussed. The sign of the parameter of each it-regressor is consistent with the economic intuition of the gravity model, though the estimate for $reer_{it}$ is not significant. If exporter's GDP increases, the exporting country produces more goods reducing its good prices and therefore increasing export to the US. If the exchange rate increases, the exporter's currency depreciates vis-a-vis the dollar and its goods become cheaper, so export to the US increases. Finally, an increase in the reer of country i means a depreciation of its currency against all currencies in the rest of the world (RoW) as a whole. Therefore, country i becomes cheaper for RoW and the demand for the products of country i increases leading to rising prices. Because country i is now more expensive for the US, its export to the US becomes smaller.

6.5 Untangling normalization

Now we apply untangling normalization to the zero-normalized time FE θ_t^0 in Figure 1. In Section 6.5.1 we look at a model without t-regressors and show the improved interpretation of the untangled time FE θ_t^u. Next, we introduce t-regressors in Section 6.5.2 and show they explain a large part of time-specific development. Finally, in Section 6.5.3 we introduce distributed lags (DL) into the model leading to our baseline specification.

6.5.1 Static model without t-regressors

Figure 2a illustrates the first main contribution of untangling normalization: the estimated FE are more informative and can be easily interpreted compared to the zero-normalized θ_t^0 in Figure 1. Figure 2a presents the estimated untangled time FE θ_t^u from Model 4 after we untangle the time FE from the constant and the common trend
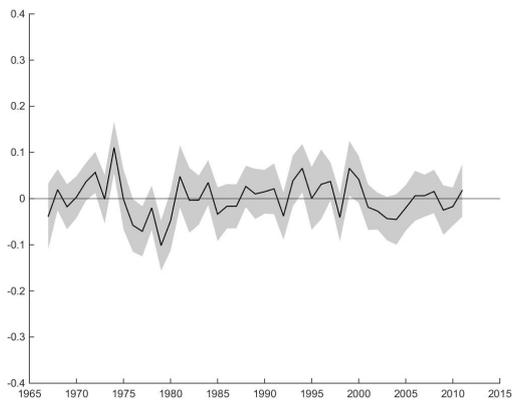
Figure 2: Untangled time FE θ_t^u: without and with t-regressors.
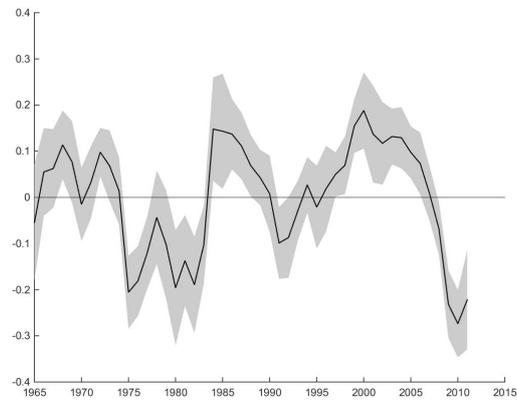
(a) Static Model 4, without t-regressors.
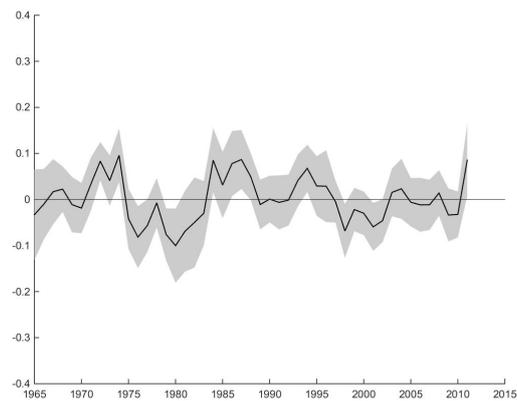
(b) Static Model 5, with t-regressors.

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

by normalizing its mean and trend to zero. We recognize a business cycle. Notice the first and the second oil crises, the dot-com bubble and, clearly visible, the financial crisis. We also see dollar exchange rate effects, such as the dollar bubble in the eighties, stimulating and then hampering exports to the US.

We perform an F-test to test the null hypothesis that all estimated θ_t^u equal zero in Model 4[8]. The null hypothesis is clearly rejected, as displayed in Table 1. The it-regressor estimates are sensitive to leaving out the time FE due to an omitted variable bias, as the comparison between Model 1 and 4 indicates.

6.5.2 Static model with t-regressors

The untangled θ_t^u in Figure 2a suggest the relevance of GDP, explaining the business cycle effect, and external value of the dollar. The theoretical gravity model suggests that US GDP, world GDP and US reer matter for export to the US. So we add these three t-regressors to the model. The traditional zero normalization does not exploit these potentially powerful regressors because it normalizes $\gamma = 0$.

Figure 2b shows the second main contribution of untangling normalization: it is capable of exploiting t-regressors. After all, it displays the untangled time FE if we include US GDP, world GDP and US reer, and the difference with 2a is striking. These three t-regressors explain most of the variation in the untangled time FE, quantified by $R^2 = 84\%$. This high explanatory power is supported by the small difference between the parameter estimates of Model 5 and Model 2 in Table 1, where in the latter model we omitted the time FE. This indicates that all estimates hardly suffer from an omitted variable bias if we omit the time FE. Hence, we can explain most of the $T = 47$ time FE by just three variables. Still, for some periods the time FE differ from zero. This is supported by the F-test in Table 1 rejecting the null hypothesis that all $\theta_t^u = 0$.

So far we have only looked at the joint contribution of the three t-regressors in explaining the estimated θ_t^u. We now evaluate their individual contributions. To examine the relevance of US GDP, we re-estimate Model 5 while leaving US GDP out as regressor. The resulting estimated θ_t^u are represented by the solid black line in Figure 3a. To make US GDP comparable with the estimated θ_t^u, we demean and detrend US GDP, and to improve visibility we rescale transformed US GDP by multiplying it with the ratio of the standard deviation of the estimated θ_t^u over the standard deviation of transformed US GDP. This gives the dashed line in Figure 3c. We see a striking re-
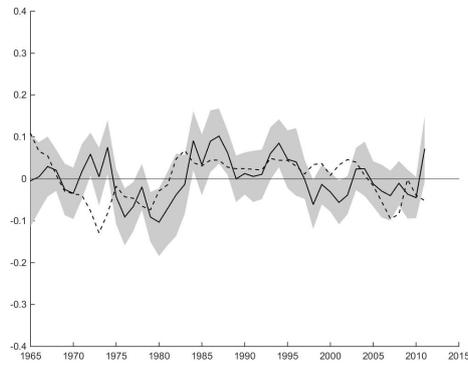
---

[8]The degrees of freedom (df) of the F-test are given by $(T - m_\theta - K_w, NT - K_d - K_x + m_d)$, where m_θ is the number of untangling normalizations in θ_t. For example, for the last column in Table 1 the df are $(34,679)$.
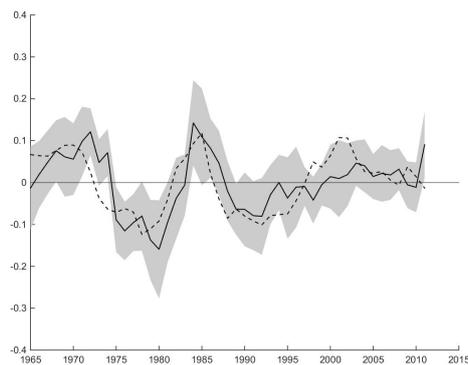
Figure 3: Untangled time FE θ_t^u: the explanatory contribution of each t-regressor.
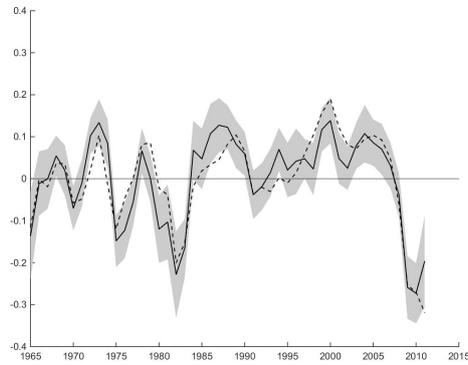
(a) Model with gdp_{Wt} and $reer_{USt}$. Dotted line: gdp_{USt}.

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

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

The subcaption shows which two t-regressors are included to estimate θ_t^u (solid line). The other t-variable is plotted (dotted line) after detrending and demeaning the t-variable and scaling it with a scaling factor, i.e. the ratio of the standard deviations of θ_t^u over that of the transformed t-variable. We multiplied world GDP and US reer with minus one (Figure 3b and 3c respectively).

semblance. The estimated θ_t^u almost exactly mimics the US business cycle. Hence, US GDP is important in explaining θ_t^u, which is supported by the high partial $R^2 = 79\%$.

In a similar way we study the relevance of world GDP and US reer, except that we multiply both by minus one for ease of comparison. Figure 3b reveals that resemblance between the estimated θ_t^u and world GDP is weak. This is supported by the low partial $R^2 = 8\%$.

Finally, in Figure 3c we look at the relevance of US reer. The estimated θ_t^u and US reer clearly exhibit the dollar swing in the eighties. The resemblance between US reer and the estimated θ_t^u is visible but not as striking as with US GDP, and the partial $R^2 = 49\%$ is therefore also lower.

6.5.3 Dynamic model with t-regressors

Until now, we have not exploited the explanatory power of lagged it- and t-regressors and the dependent variable itself. Including it-regressors may take some noise out of the estimated θ_t^u, and including t-regressors may help to explain the θ_t^u. Hence, we introduce dynamics in (18).

There is economic motivation for adding dynamics to an empirical gravity model. Often traders write down contracts in the previous period to export goods this period and often also for the future. There can also be dynamics in export itself. Export in the previous period might be important for explaining export in the current period by for instance habit formation (see for instance Mayer (1980) and Vatan (2013)).

One can capture dynamics with an autoregressive distributed lag (ADL) model or with a DL model with autoregressive errors. Both models are very similar. We use a DL model, because in an ADL framework part of the dynamics in the time FE will be taken over by lagged export terms, even though this dynamics might actually be driven by time-specific developments. Moving such dynamics to the lagged export term hampers the interpretation of the untangled time FE. The DL approach is better able to separate the sources of dynamics. Finally, an ADL model does not affect our finding that time-specific developments can be explained by US GDP, US reer and world GDP. Again we opt for adjusting the standard errors following Newey & West (1987), instead of modeling the serial correlation in the error term explicitly.

We add two lags in our DL model, in the form of the first difference for each independent variable and its lag, for example Δgdp_{it} and $\Delta gdp_{i,t-1}$. In this way the parameters of the level variables capture the long-run effects and are comparable to the static model parameters. Including more lag terms is not necessary and leads to the same main conclusion. We only report and discuss the DL long-run effect estimates.

Figure 2c shows the estimated θ_t^u for the DL model with t-regressors. They are close to zero. This is confirmed by the F-test, which cannot reject the null hypothesis that all $\theta_t^u = 0$. We find that the t-regressors explain the estimated θ_t^u almost completely, with $R^2 = 91\%$. If we adjust for the large number of DL parameters we find the adjusted $R^2 = 88\%$.

Model 6 Table 1 displays the underlying estimation results. Almost all the signs in Model 6 are in line with the gravity model. The β estimates in Model 6 all have the correct sign and are comparable to Model 4, as discussed in Section 6.4. We here focus on the γ estimates. US GDP has a positive effect on exports to the US because an increase in US GDP means that the US has more income and is able to import more from country i. World GDP has a negative impact because of two reasons. First, if world GDP increases, the demand for the goods of country i by RoW increases leading to a higher price for its goods. This reduces export to the US. Second, RoW produce more goods making them cheaper and therefore the US substitutes their import from country i to RoW. The final variable, US reer, has a negative effect, while the gravity model predicts a positive sign. After all, when US reer increases, that is, a depreciation of the dollar against all currencies in the RoW, it becomes more expensive to import from the RoW by the US and as a consequence the US demands more products from country i and export goes up.

If we omit time FE, the estimates for Model 3 reveal that this hardly influences the estimates of both the it- and t-regressors, so we can safely leave out θ_t. This supports the conclusion based on the F-test and Figure 2c that all θ_t^u are not significantly different from zero.

The γ estimates are conditional on our normalization, so they reflect the pseudo-true values of the impact of the t-regressors, not the true values. The differences becomes smaller if the remaining time effects θ_t^u, what is left after exploiting the t-regressors, are closer to zero; see Section 4. As we find that the remaining time FE are redundant, we conclude that the γ estimates reflect the true values of the impact of the t-regressors.

In summary, the application has illustrated two advantages of untangling normalization. First, the estimated untangled time FE θ_t^u are easier to interpret than the estimated zero-normalized time FE θ_t^0 and using this property we can identify potential t-regressors that might be important in explaining the time-specific developments. Second, we can estimate the parameters of the t-regressors and time FE simultaneously and determine how important the former are in explaining the later. We can explain the $T = 47$ time FE for a large part by only three t-regressors. If we include two DL-terms of each t-regressors we explain up to 90% of the time-specific developments making the

remaining θ_t^u redundant. We thus obtain estimates for t-regressors that reflect their true value, despite the fact that estimating such impacts has been a notorious problem in the literature.

6.6 Poisson pseudo maximum likelihood

Instead of the LSDV estimation method we used so far, one could also use Poisson pseudo maximum likelihood (PPML) estimation, as proposed by Santos-Silva & Tenreyro (2006), to estimate a gravity model. Using the notation of (2), LSDV is based on the assumption that $\mathbb{E}\{\log(y)|D, Z, X\}$ is a linear function of the regressors, which include the fixed effects and constant regressors. PPML is used if one prefers assuming that $\log(\mathbb{E}\{y|D, Z, X\})$ is a linear function of the regressors. In both cases the fixed effects and the constant regressors enter in the linear part, and that implies that our approach can be applied in combination with both estimation methods[9].

Figure 4 displays the estimated untangled time FE θ_t^u using PPML. We can see that they are small and in most periods insignificant. As for LSDV, we cannot reject the null hypothesis $\theta_t^u = 0$ for the PPML estimates.

In Model 7 Table 2 we display the estimation results for PPML. Most estimates are in line with the LSDV estimates, except for the exchange rate estimates. Explaining the latter goes beyond the focus of this paper.

6.7 Fixed effects vector decomposition

An alternative to the untangling approach is the fixed effects vector decomposition (FEVD), which consist of two stages. The first stage regression is identical to the (only) regression we have and one typically uses LSDV, as we do here. In the second stage, FEVD uses the estimated zero-normalized time FE from the first stage as dependent variable and regresses it on the t-regressors:

$$\widehat{\theta}_t^0 = \alpha + \tau \cdot t + w_t \gamma + \eta_t, \tag{19}$$

where η_t is the error term in this second stage regression. In Section 4.4 we have compared FEVD to our approach and argued why we prefer the latter. We now illustrate

---

[9]PPML uses export in levels as dependent variable, that is, $\exp(exp_{iUSt})$. Because the latter is expressed in real currency of country i, countries with a high nominal exchange rate (for example Japan: about 100 yen for one dollar) will have larger absolute error terms and play a larger role in the pseudo log-likelihood and therefore dominate the estimation results. To avoid this scale dependency, we transform the dependent variable into US dollars using the purchasing power parity (PPP) exchange rate of the base year. Note that this rescaling does not matter for LSDV, because there the log of the PPP rate ends up in α_i.

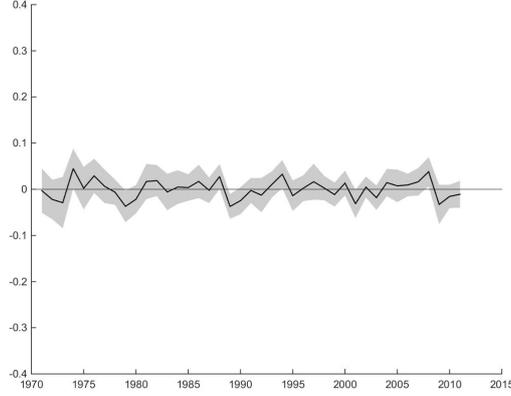Figure 4: Untangled time FE θ_t^u: Model 7 with t-regressors estimated by PPML.

Table 2: Estimation results for exp_{iUSt} using different estimators based on (18).

Estimation method	LSDV	PPML	FEVD		
2nd stage reg. incl.	–	–	α, τ	No α, τ	α; no τ
Model	6	7	8	8.1	8.2
gdp_{it}	1.40*	1.55*	1.40*	...	...
	(0.20)	(0.20)	(0.20)		
rer_{iUSt}	0.71*	2.28*	0.71*	...	...
	(0.22)	(0.26)	(0.22)		
$reer_{it}$	-0.43	-2.01*	-0.43	...	...
	(0.25)	(0.27)	(0.25)		
gdp_{USt}	2.25*	3.07*	2.25*	5.92*	2.42*
	0.39	(0.77)	(0.16)	(0.46)	(0.20)
gdp_{Wt}	-1.34*	-2.64*	-1.34*	-4.40*	-1.03*
	0.47	(0.57)	(0.44)	(0.37)	(0.16)
$reer_{USt}$	-0.54*	1.31*	-0.54*	0.43	-0.55*
	(0.27)	(0.26)	(0.09)	(0.29)	(0.10)
F-/Wald-test $\theta_t^u = 0$	1.26	24.33	–	–	–
p-value	0.15	0.14	–	–	–

All models use a DL model with 2 lags for every regressor. Model 7 uses a DL model with 6 lags for every regressor. In case of PPML we apply a Wald-test with 18 df instead of an F-test, because the F-statistic is a monotone transformation of the likelihood ratio statistic and the latter is invalid in PPML because the information matrix equality does not hold. In the second stage of the FEVD we use (19), where we in- or exclude α and τ. The first stage in all displayed FEVD cases is identical as represented by the dots. Standard errors are between brackets. * indicates significance at 5% level.

that using our application.

Model 8 Table 2 displays the FEVD results. If we apply the second stage regression (19), we find γ estimates that are identical to the untangling normalization approach, reflecting that both approaches use the same orthogonality relations (FEVD treats them as restrictions, we only as normalizations). The residuals $\widehat{\eta}_t$ are thus the same as our $\widehat{\theta}_t^u$, so Figure 2c visualizes the former as well. However, the standard errors for all three γ estimates differ substantially, because those of FEVD do not account for two-step nature of FEVD. Untangling normalization improves on this, because it has no second estimation step.

So far, we have taken the variables on the RHS and LHS of (19) for granted. As discussed in Section 4.4, our choice is not obvious. We start with the RHS. It is not immediately clear what to include besides the w_t regressors, and this decision can have a big impact on the estimated γ parameters. Plümper & Troeger (2007) only include w_t in the second stage. Pesaran & Zhou (2014) show, however, that it is important to include a constant on the RHS to prevent biased γ estimates. Indeed, if we estimate (19) without α and τ, see Model 8.1 in Table 2, we find severely biased estimates and also the standard errors are affected.

We extend this finding by considering the overall trend. If we include α but omit τ in (19), we again find altered estimates and standard errors, see Model 8.2 in Table 2. Although the differences are smaller than for the case where also the constant α is left out, omitting the overall trend τ can also lead to biases.

But how do we know whether to include a constant and/or overall trend in the second stage? Plümper & Troeger (2007) apparently think that leaving out the constant is warranted, but Pesaran & Zhou (2014) disagree. Perhaps the plot of $\widehat{\theta}_t^0$ helps. Figure 1 shows the plot for two different zero normalizations (strictly speaking the figure is based on the static model, but the figure for the DL model that we need is very similar). The $\widehat{\theta}_t^0$ series heavily depend on the specific zero normalization that is used. If we normalize $\alpha = \tau = 0$, it is tempting to conclude from the confidence band in Figure 1a that including a constant matters, but that one can safely leave out a trend. If we instead normalize $\theta_{T-1} = \theta_T = 0$, then Figure 1b suggests that both a constant and trend may be left out or, if one stresses the significant value at the end of the sample, both may matter. We have shown that including both is the right choice, but the point is that it is easy to choose incorrect specifications in the second stage of the FEVD.

Also on the LHS we have to make choices which variables to include. Should we include $\widehat{\alpha}_i^0$ and $\widehat{\tau}_i^0 \cdot t$, knowing that they interfere with $\widehat{\theta}_t^0$? The similarity of the

28

estimates between Model 6 and 8 in Table 2 suggest that the answer is negative, as long as one is not interested in the estimates of α and τ. However, this again shows that FEVD demands a number of choices from the empirical researcher. Untangling normalization avoids such choices, because it transforms all estimated fixed effects and constant regressors in a unique way.

6.8 Sensitivity analysis

We have performed several sensitivity checks to confirm that our results are robust for different model specifications.

6.8.1 No country-specific trends

Table 3: Sensitivity regarding trends and cointegration, based on DL version of (18).

		No $\tau_i \cdot t$	DOLS
Model	6	9	10
gdp_{it}	1.40*	1.94*	1.58*
	(0.20)	(0.24)	(0.21)
rer_{iUSt}	0.71*	0.39	0.82*
	(0.22)	(0.28)	(0.29)
$reer_{it}$	-0.43	0.12	-0.69*
	(0.25)	(0.28)	(0.33)
gdp_{USt}	2.25*	1.59*	2.09*
	(0.39)	(0.54)	(0.33)
gdp_{Wt}	-1.34*	-1.97*	-1.43*
	(0.47)	(0.69)	(0.43)
$reer_{USt}$	-0.54*	-0.84*	-0.46*
	(0.27)	(0.34)	(0.09)
F-test $\theta_t^u = 0$	1.26	0.60	1.29
p-value	0.15	0.96	0.13

Model 10 explicitly accounts for cointegration between exp_{iUSt}, gdp_{it}, rer_{iUSt} and $reer_{it}$ and uses DOLS estimation with two leads and lags for all non-stationary regressors. Standard errors are between brackets. * indicates significance at 5% level.

The number of papers that include trend FE into a gravity type of model is growing, for exampleBun & Klaassen (2007) and Baier *et al.* (2014). We show that the inclusion of $\tau_i \cdot t$ is not driving our main results, but that it matters for the estimates.

In Model 9 Table 3 we display the estimation results for a model without $\tau_i \cdot t$. The parameters of the t-regressors still have a similar size compared to Model 6 and are able to capture most of the time effects. This is supported by the F-test, which does not reject the null hypothesis $\theta_t^u = 0$ for all t. This confirms that $\tau_i \cdot t$ is not driving our main result, that we can explain the time FE by three variables US GDP, world

GDP, and US reer.

We see that in Model 9 the gdp_{it} parameter is substantially larger than its counterpart in Model 6. The intuition is that, if we omit $\tau_i \cdot t$, we do not allow for country trends, and because the gdp_{it} variable is the only it-regressor with a clear trend, it will try to fit the uncaptured country trends in export. This leads to a bias of the estimated gdp_{it} parameter. At the same time, the country trends show up in the unexplained part, so that the exchange rate impact becomes more difficult to find. Finally, the null hypothesis $\tau_i = 0$ for all i is clearly rejected, with a p-value of 0.00. So it is advisable to include $\tau_i \cdot t$ in the model.

6.8.2 Non-stationarity and cointegration

In previous sections we have pretended that all variables in (18) are stationary. However, the variables exp_{iUSt} and gdp_{it} are most likely non-stationary, and rer_{iUSt} and $reer_{it}$ may also be (close to) non-stationary. We show that, even if one accounts for non-stationarity issues, our results remain valid.

We first test the null hypothesis that the variables under consideration have a unit root for all countries, using the four Fisher (1932) type test statistics proposed by Choi (2001) and available in STATA. We account for lags, time effects, parameter heterogeneity, and for all series we add a drift term to the test equation, and for export and gpd we also add a trend. Not surprisingly, we conclude that all four variables have a unit root.

Next, we apply the panel cointegration test of Pedroni (1999). We allow for time effects, country trends, and parameter heterogeneity in the test equation. There is strong evidence for a cointegration relation between exp_{iUSt} and gdp_{it}, but at first we find no evidence for cointegration if we add rer_{iUSt} and $reer_{it}$. This may be due to the small N in our data. If we include all bilateral combinations between the 17 countries and the US, increasing N from 17 to 306, we find some evidence for cointegration between the four variables. We thus assume that there exists such a cointegration relation.

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 & Sul (2003). First, we estimate the cointegrating regression of exp_{iUSt} on gdp_{it}, rer_{iUSt} and $reer_{it}$, adding two leads and lags of the first differences of the regressors (combinations of 0-3 leads and lags yield similar results), allowing its coefficients to be heterogeneous for all countries, and including the full set of fixed effects. Second, we substitute the estimated β_1, β_2 and β_3 into (18) with

DL terms and estimate the remaining parameters using LSDV and apply untangling normalization.

Model 10 in Table 3 displays the result. They do not differ much from the baseline results in Model 6. The estimated θ_t^u are close to zero, comparable to Figure 2c. We again cannot reject the null hypothesis $\theta_t^u = 0$.[10] Hence, we conclude that accounting for non-stationarity and cointegration does not alter our main conclusion.

7 Conclusion

We have developed a new normalization method: untangling normalization. It can be used to disentangle different FE-types from each other, easing their interpretation substantially. We can also apply it to untangle constant regressors from FE spanning the same dimensions by orthogonalizing the latter with respect to the former, enabling us to simultaneously estimate both the parameters of the constant regressors and the FE, thus exploiting the explanatory power of the constant regressors. This is not possible in standard normalization methods like zero normalization. The untangled parameter estimates and standard errors are just linear transformations from its zero-normalized counterparts and can be obtained without the need for a new estimation.

We have applied untangling normalization to a simple gravity-based trade model, where we model export of 17 OECD countries to the US. This illustrates the two advantages of untangling normalization. First, the estimated untangled time FE are much easier to interpret than the estimated zero-normalized time FE. The estimated untangled time FE clearly show a business cycle pattern together with a dollar exchange rate effect. Using these insights together with gravity theory, we obtain three t-regressors, US GDP, world GDP and US reer, that might be important in explaining the time-specific developments. Second, we can estimate the impact of the t-regressors and time FE and determine how important the t-regressors are. We can explain the 47 time FE for 90% by only three t-regressors making the remaining estimated untangled time FE redundant. Hence, we have estimates of the true values of the impacts of the t-regressors, even though that is typically considered to be beyond reach. All these results could not have been obtained using zero normalization.

The current paper is about two dimensional panels, which have dimension it. One

---

[10]The t-variables gdp_{USt}, gdp_{Wt} and $reer_{USt}$ are non-stationary. Although the time dummies themselves are deterministic, the estimated series of time effects need not be stationary. Figure 2c suggests that the series of estimated time effects and the t-regressors are cointegrated. Hence, the impacts γ of the t-regressors are consistently estimated and only the standard errors might be invalid. But the standard errors will probably be good approximations given the small changes in the standard errors of the it-regressors between LSDV and DOLS. Hence, we use standard inference for the estimated γ.

idea for future work is to extend our untangling normalization method to a three dimensional panel setting. Untangling normalization can then be applied to bilateral trade flows.

References

Anderson, James E., & Wincoop, Eric van. 2003. Gravity with Gravitas: A Solution to the Border Puzzle. *The American Economic Review*, **93**(1), pp. 170–192.

Baier, Scott L., Bergstrand, Jeffrey H., & Feng, Michael. 2014. Economic integration agreements and the margins of international trade. *Journal of International Economics*, **93**(2), 339 – 350.

Breusch, Trevor, Ward, Michael B., Nguyen, Hoa Thi Minh, & Kompas, Tom. 2011a. FEVD: Just IV or Just Mistaken? *Political Analysis*, **19**(2), 165–169.

Breusch, Trevor, Ward, Michael B., Nguyen, Hoa Thi Minh, & Kompas, Tom. 2011b. On the Fixed-Effects Vector Decomposition. *Political Analysis*, **19**(2), 123–134.

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

Choi, In. 2001. Unit root tests for panel data. *Journal of International Money and Finance*, **20**(2), 249–272.

Fisher, Ronald Aylmer Sir. 1932. *Statistical methods for research workers*. Fourth ed. - revised and enlarged edn. Edinburgh Oliver and Boyd. "Sources used for data and methods": p. 297-299. Bibliography: p. 300-304.

Greene, William. 2011. Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time-Invariant Variables in Fixed Effects Models? *Political Analysis*, **19**(2), 135–146.

Hausman, Jerry A., & Taylor, William E. 1981. Panel Data and Unobservable Individual Effects. *Econometrica*, **49**(6), 1377–1398.

Klaassen, Franc, & Teulings, Rutger. 2015. **Working paper:** *Untangling Multilateral Boundaries*. Tech. rept. University of Amsterdam.

Mark, Nelson C., & Sul, Donggyu. 2003. Cointegration Vector Estimation by Panel DOLS and Long-run Money Demand. *Oxford Bulletin of Economics and Statistics*, **65**(5), 655–680.

Mayer, Wolfgang. 1980. Two implications of habit formation for international trade. *Economics Letters*, **6**(4), 367 – 371.

Mundlak, Yair. 1978. On the Pooling of Time Series and Cross Section Data. *Econometrica*, **46**(1), 69–85.

Newey, Whitney K, & West, Kenneth D. 1987. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, **55**(3), 703–08.

Newey, Whitney K, & West, Kenneth D. 1994. Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, **61**(4), 631–53.

Pedroni, Peter. 1999. Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors. *Oxford Bulletin of Economics and Statistics*, **61**(0), 653–70.

Pesaran, M. Hashem, & Zhou, Qiankun. 2014 (September 5). *Estimation of Time-Invariant Effects in Static Panel Data Models*. Research Paper 14.08. CAFE.

Plümper, Thomas, & Troeger, Vera E. 2007. Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects. *Political Analysis*, **15**(2), 124–139.

Santos-Silva, J. M. C., & Tenreyro, S. 2006. The Log of Gravity. *The Review of Economics and Statistics*, **88**(4), 641–658.

Suits, Daniel B. 1984. Dummy Variables: Mechanics v. Interpretation. *The Review of Economics and Statistics*, **66**(1), 177–80.

Vatan, Antoine. 2013. Habit formation and strategic interactions: A new gain from trade? *Economics Letters*, **121**(1), 101–104.