Identifying Directional Challenges in Gravity Equations with Balanced Dyadic Panel Data *

Stefan Smutny

University of Vienna

April 8, 2024

Abstract

Gravity equations are widely used in international trade and migration studies to estimate the flow of goods, services, and people between countries. However, existing tools for estimating gravity equations lack the capability to account for directional effects accurately. This brief note addresses this gap by formally demonstrating that utilizing a full set of fixed effects to incorporate multilateral resistance yields identical estimates for both outflow and inflow dependent variables in gravity models.

Keywords: Gravity Equations, Fixed Effects, Panel Data Models, Directional Effects

JEL Codes: C23, F14, F22, E65

*Stefan Smutny: Department of Economics, University of Vienna, Oskar Morgenstern Platz 1, 1090 Vienna, stefan.smutny@univie.ac.at. .

1 Brief Introduction

Gravity equations are consistently estimated using a comprehensive set of fixed effects to control for multilateral resistances, as highlighted by [Weidner and Zylkin, 2021]. This approach effectively absorbs state-specific effects, allowing for a causal interpretation of estimation results. However, a critical challenge in interpreting these results has often been overlooked. The estimated coefficients capture changes in joint flows of the dependent variable (such as trade or migration), rather than distinguishing directional effects. This limitation arises from the utilization of fixed effects and within-transformations of all variables in the specification. Consequently, once these fixed effects are integrated, and the underlying dyadic panel is perfectly balanced, policy variables will exert the same effect on both inflows and outflows. This note formally demonstrates that distinguishing effects on in- and outflows is not achievable with the prevailing techniques in the gravity literature. Thus, caution is warranted when interpreting gravity effects directionally, as separate estimates for in- and outflows emerge solely from an unbalanced panel rather than the distinction of these directions.

2 Statement of the Problem

To show the directional symmetry problem of gravity estimation in a tractable context, we can start by stating a simple gravity equation of flows between i and j such that

$$y_{i,j} = \exp[\beta x_{i,j} + \delta_i + \delta_j]\eta_{i,j}, \tag{1}$$

where $x_{i,j}$ defines the policy variable of interest, δ_i and δ_j the origin and destination fixed effects. In order to account for these fixed effects, we create within transformations of our variables. Since equation (1) applies two fixed effects, the required transformation has to account for both group averages, i.e., double demeaning. For any variable $a_{i,j}$, we can define the within transformation as $\tilde{a}_{i,j} = a_{i,j} - [\bar{a}_{i,\cdot} + \bar{a}_{\cdot,j}] + \bar{\bar{a}}$, where $\bar{a}_{i,\cdot} = J^{-1}\sum_j a_{i,j}$ and $\bar{\bar{a}} = [I \times J]^{-1}\sum_i \sum_j a_{i,t}$. Transforming variables in equation (1) accordingly leaves us with

$$\tilde{y}_{i,j} = \exp[\beta \tilde{x}_{i,j}]\tilde{\eta}_{i,j}, \tag{2}$$

where the model is either estimated in its multiplicative form through a GLM procedure such as FE-PPML, or by log-linearizing and using OLS. With the latter specification, we can show that the within transformation required to account for the state specific fixed effects lead to the following observation:

Lemma 1 *For any pair-wise varying dummy it holds that* $\tilde{x}_{i,j} = \tilde{x}_{j,i}$.

The within transformation yields a demeaned variable that now only holds pairwise information. Consequently, we deprive our policy variable of state specific information. Hence, the transformed policy variable defines whether a given $i \times j$-pair has a different policy regime, without providing information on which of the two states has implemented the policy. Recall that $y^{out}_{i,j} = y^{in}_{j,i}$, which then implies that also $\tilde{y}^{out}_{i,j} = \tilde{y}^{in}_{j,i}$ we can follow:

Corollary 1.1 *If Lemma 1 holds, we estimate the same specification for* $y^{out}_{i,j}$ *and* $y^{in}_{i,j}$ *and therefore obtain the same* $\hat{\beta}$ *for both dependent variables.*

This finding leaves us with interpretative challenges. Rather than estimating the effects of the tax policy on in- and out-migration, we estimate the effect of either one state of the $i \times j$-pair implementing the policy on the pair's migration flow.

3 Proof

Proof of Lemma 1: Assume a dummy variable $x_{i,j} = z_i m_{i,j} \in [0,1]$, where $z_i, m_{i,j} \in [0,1]$. This implies that $I^{-1}\sum_i x_{i,j}, J^{-1}\sum_j x_{i,j} \in [0,1]$. Further, assume that $I = J$. Hence, $\bar{x}_{i,\cdot}, \bar{x}_{\cdot,j} \in [0,1]$. Show that

$$x_{i,j} - [\bar{x}_{i,\cdot} + \bar{x}_{\cdot,j}] = x_{j,i} - [\bar{x}_{j,\cdot} + \bar{x}_{\cdot,i}].$$

There are several cases to consider. Since our dummy variable is pairwise varying, we have four different possible combinations of $x_{i,j}$ and $x_{j,i}$.[1]

Case 1: $x_{i,j} = 1$, $x_{j,i} = 0$ where $z_i = 1, z_j = 0$, $m_{i,j} = 1$.

Here, we have that $\tilde{x}_{i,j} = 1 - [\bar{x}_{i,\cdot} + \bar{x}_{\cdot,j}] + \bar{\bar{x}}$ and $\tilde{x}_{j,i} = 0 - [\bar{x}_{j,\cdot} + \bar{x}_{\cdot,i}] + \bar{\bar{x}}$. First, we note that since $x_{j,i} = 0 \,\forall\, i$ we have $\bar{x}_{j,\cdot} = 0$. Similarly, since $x_{i,j} = 1 \,\forall\, j$ we have $\bar{x}_{i,\cdot} = 1$. As $m_{i,j} = 1 \,\forall\, i, j$, we further have $\bar{x}_{\cdot,j} = \bar{x}_{\cdot,i} \,\forall\, i, j$. Therefore, we can write

$$\begin{aligned}\tilde{x}_{i,j} &= 1 - [1 + \bar{x}_{\cdot,i}] + \bar{\bar{x}} \\ &= 0 - [0 + \bar{x}_{\cdot,i}] + \bar{\bar{x}} = \tilde{x}_{j,i}\end{aligned}$$

Case 2: $x_{i,j} = 0$, $x_{j,i} = 1$ where $z_i = 1, z_j = 0$, $m_{i,j} = 1$.

We can show that $\tilde{x}_{i,j} = \tilde{x}_{j,i}$, analogously to the case above, with i, j flipped.

Case 3: $x_{i,j} = 1$, $x_{j,i} = 1$ where $z_i = 1, z_j = 1$, $m_{i,j} = 1$.

First, we note that $x_{i,j} = x_{j,i} = 1 \,\forall\, i, j$. Therefore, we have $\bar{x}_{i,\cdot} = \bar{x}_{j,\cdot} = 1 \,\forall\, i, j$. Analogously to before, as $m_{i,j} = 1 \,\forall\, i, j$, we further have $\bar{x}_{\cdot,j} = \bar{x}_{\cdot,i} \,\forall\, i, j$. Therefore, we can state

$$\tilde{x}_{i,j} = 1 - [1 + \bar{x}_{\cdot,i}] + \bar{\bar{x}} = 1 - [1 + \bar{x}_{\cdot,i}] + \bar{\bar{x}} = \tilde{x}_{j,i}$$

Case 4: $x_{i,j} = 0$, $x_{j,i} = 0$ where $z_i = 0, z_j = 0$, $m_{i,j} = 1$.

Again, first, we note that $x_{i,j} = x_{j,i} = 0 \,\forall\, i, j$. Therefore, we have $\bar{x}_{i,\cdot} = \bar{x}_{j,\cdot} = 0 \,\forall\, i, j$. And again, as $m_{i,j} = 1 \,\forall\, i, j$, we further have $\bar{x}_{\cdot,j} = \bar{x}_{\cdot,i} \,\forall\, i, j$. Similarly to above, we have

$$\tilde{x}_{i,j} = 0 - [0 + \bar{x}_{\cdot,i}] + \bar{\bar{x}} = 0 - [0 + \bar{x}_{\cdot,i}] + \bar{\bar{x}} = \tilde{x}_{j,i}$$

Case 5: $m_{i,j} = 0$, therefore $x_{i,j} = x_{j,i} = x_{i,i}$.

It follows, immediately, that $\tilde{x}_{i,j} = \tilde{x}_{i,i} = \tilde{x}_{j,i}$. ■

[1] The proof is very straight forward, as the pairwise moving property does not come from a definition of adoption differentials but rather from the interstate migration dummy $m_{i,j}$.

Note that the simplicity of this proof comes from the fact that $x_{i,j,t}$ only varies in j due to $m_{i,j}$. If we however were to define a different dummy $x^*_{i,j,t}$ that varies in j by definition, this proof becomes a bit more involved. Let us define $x^*_{i,j,t} = 1$ iff $z_i > z_j$. In this case, we design the bilateral adoption dummy as a non-linear combination of unilateral policy variables z_i and z_j. The main reason we do not use such a dummy in our main specification are its implicit assumptions. It must hold that $z_i = z_j = 0$ and $z_i = z_j = 1$ both imply $x^*_{i,j,t} = 0$. Note that $z_i = 0, z_j = 1$ also implies $x^*_{i,j,t} = 0$. Hence, the interpretation of such a variable becomes challenging. Nonetheless, we can show, that for $x^*_{i,j,t}$ the same Lemma 1 holds.

Proof of Lemma 1 for $x^*_{i,j,t}$: Assume a dummy variable $x^*_{i,j} = \{0, 1\}$ such that $I^{-1}\sum_i x^*_{i,j}, J^{-1}\sum_j x^*_{i,j} \in [0, 1]$. Further, assume that $I = J$. Hence, $\bar{x^*}_{i,\cdot}, \bar{x^*}_{\cdot,j} \in [0, 1]$. Show that

$$x^*_{i,j} - [\bar{x^*}_{i,\cdot} + \bar{x^*}_{\cdot,j}] = x^*_{j,i} - [\bar{x^*}_{j,\cdot} + \bar{x^*}_{\cdot,i}].$$

There are several cases to consider. Since our dummy variable is pairwise varying, we have four different possible combinations of $x^*_{i,j}$ and $x^*_{j,i}$.

Case 1: $x^*_{i,j} = 1, x^*_{j,i} = 0$ where $z_i = 1, z_j = 0$.
Here, we have that $\tilde{x}^*_{i,j} = 1 - [\bar{x}^*_{i,\cdot} + \bar{x}^*_{\cdot,j}] + \bar{\bar{x}}^*$ and $\tilde{x}^*_{j,i} = 0 - [\bar{x}^*_{j,\cdot} + \bar{x}^*_{\cdot,i}] + \bar{\bar{x}}^*$. First, we note that since $x^*_{j,i} = 0 \; \forall \; i$ we have $\bar{x}^*_{j,\cdot} = 0$. Similarly, as $x^*_{j,i} = 0 \; \forall \; j$ we have $\bar{x}^*_{\cdot,i} = 0$.[2] It follows directly that $\tilde{x}^*_{j,i} = 0 - [\bar{x}^*_{j,\cdot} + \bar{x}^*_{\cdot,i}] + \bar{\bar{x}}^* = \bar{\bar{x}}^*$.

Next, we notice that, by design of our dummy, we have $\bar{x}^*_{\cdot,j} = 1 - \bar{x}^*_{i,\cdot}$ where we must consider three different cases.

Case 1.1: $\bar{x}_{i,\cdot}, \bar{x}_{\cdot,j} < 1$. Then, we have $\tilde{x}^*_{i,j} = 1 - [\bar{x}^*_{i,\cdot} + (1 - \bar{x}^*_{i,\cdot})] + \bar{\bar{x}}^* = \bar{\bar{x}}^*$. Notice, that $\bar{x}^*_{i,\cdot} > \bar{x}^*_{\cdot,j}$ iff $\sum_i z_i > \sum_j z_j$.

[2]Note that this is only possible due to the dummy's specific structure where $x^*_{i,j} = 0$ if $z_i = z_j = 1$ or $z_i = z_j = 0$

Case 1.2: $\bar{x}^*_{i,\cdot} = 1$ where $\sum_i z_i = 1$.[3] Since $\bar{x}^*_{\cdot,j} = 1 - \bar{x}^*_{i,\cdot} = 0$, we have $\tilde{x}^*_{i,j} = 1 - [\bar{x}^*_{i,\cdot} + 0] + \bar{\bar{x}}^* = \bar{\bar{x}}^*$.

Case 1.3: $\bar{x}^*_{\cdot,j} = 1$ where $\sum_j z_j = 1$.[4] Since $\bar{x}^*_{\cdot,j} = 1 - \bar{x}^*_{i,\cdot} = 1$ gives us that $\bar{x}^*_{i,\cdot} = 0$, we have $\tilde{x}^*_{i,j} = 1 - [0 + (1 - 0)] + \bar{\bar{x}}^* = \bar{\bar{x}}^*$.

Therefore, we can write $\tilde{x}^*_{i,j} = \tilde{x}^*_{j,i} = \bar{\bar{x}}^*$ with $x^*_{i,j} - [\bar{x}^*_{i,\cdot} + \bar{x}^*_{\cdot,j}] = x^*_{j,i} - [\bar{x}^*_{j,\cdot} + \bar{x}^*_{\cdot,i}] = 0$.

Case 2: $x^*_{i,j} = 0$, $x^*_{j,i} = 1$ where $z_i = 0, z_j = 1$.

This case follows the same argument as the one above. We have $\tilde{x}^*_{i,j} = 0 - [\bar{x}^*_{i,\cdot} + \bar{x}^*_{\cdot,j}] + \bar{\bar{x}}^* = \bar{\bar{x}}^*$ as $x^*_{i,j} = 0 \,\forall\, i, j$. We also have $\bar{x}^*_{\cdot,i} = 1 - \bar{x}^*_{j,\cdot}$ and therefore $\tilde{x}^*_{j,i} = 1 - [\bar{x}^*_{j,\cdot} + (1 - \bar{x}^*_{j,\cdot})] + \bar{\bar{x}}^* = \bar{\bar{x}}^*$.

Therefore, we can again write $\tilde{x}^*_{i,j} = \tilde{x}^*_{j,i} = \bar{\bar{x}}^*$ with $x^*_{i,j} - [\bar{x}^*_{i,\cdot} + \bar{x}^*_{\cdot,j}] = x^*_{j,i} - [\bar{x}^*_{j,\cdot} + \bar{x}^*_{\cdot,i}] = 0$.

Case 3: $x^*_{i,j} = 0$, $x^*_{j,i} = 0$ where $z_i = z_j = 0$.

First, we note that $x^*_{i,j} = x^*_{j,i} = 0 \,\forall\, i, j$. Therefore, we have $\bar{x}^*_{i,\cdot} = \bar{x}^*_{j,\cdot} = 0$. Consider $k, l \neq i, j$ and recall that $\mathbb{1}_{x^*_{k,l}}(z_k > z_l)$. Hence, if $\exists\, k$ s.t. $z_k > z_l$, we have $\bar{x}^*_{\cdot,i} > 0 \,\forall\, i$. Therefore, we have $\bar{x}^*_{\cdot,i} = \bar{x}^*_{\cdot,j} \geq 0$. It follows that $\tilde{x}^*_{i,j} = 0 - [0 + \bar{T}_{\cdot,j}] + \bar{\bar{x}}^* = \bar{x}^*_{\cdot,j} + \bar{\bar{x}}^* \,\forall\, j$.

Therefore, we can write $\tilde{x}^*_{i,j} = \tilde{x}^*_{j,i} = \bar{x}^*_{\cdot,i} + \bar{\bar{x}}^* = \bar{x}^*_{\cdot,j} + \bar{\bar{x}}^*$ with $x^*_{i,j} - [\bar{x}^*_{i,\cdot} + \bar{x}^*_{\cdot,j}] = x^*_{j,i} - [\bar{x}^*_{j,\cdot} + \bar{x}^*_{\cdot,i}] = 0$.

Case 4: $x^*_{i,j} = 0$, $x^*_{j,i} = 0$ where $z_i = z_j = 1$.

Similar to the case above, we note that $x^*_{i,j} = x^*_{j,i} = 0 \,\forall\, i, j$. Therefore, we have $\bar{x}^*_{\cdot,j} = \bar{x}^*_{\cdot,i} = 0$ (as already shown in Case 2). Further, note that, as $z_i = 1$, $J^{-1} \sum_j x^*_{i,j} \geq 0 \,\forall\, j$. Hence $\bar{x}^*_{i,\cdot} = \bar{x}^*_{j,\cdot} \geq 0$. It

[3]The extreme case represents a situation where all states adopt the tax.
[4]The extreme case represents a situation where no state adopts the tax.

follows that $\tilde{x}^*_{i,j} = 0 - [\bar{x}^*_{i,\cdot} + 0] + \bar{\bar{x}}^* = \bar{x}^*_{i,\cdot} + \bar{\bar{x}}^* \; \forall \; i$.

Therefore, we can write $\tilde{x}^*_{i,j} = \tilde{x}^*_{j,i} = \bar{x}^*_{i,\cdot} + \bar{\bar{x}}^* = \bar{x}^*_{j,\cdot} + \bar{\bar{x}}^*$ with $x^*_{i,j} - [\bar{x}^*_{i,\cdot} + \bar{x}^*_{\cdot,j}] = x^*_{j,i} - [\bar{x}^*_{j,\cdot} + \bar{x}^*_{\cdot,i}] = 0$.

■

References

Weidner, M. and T. Zylkin (2021). Bias and Consistency in Three-way Gravity Models. *Journal of International Economics 132*, 103513.