DIW BERLIN

1748

# Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2018

# Interactions in Fixed Effects Regression Models

Marco Giesselmann and Alexander Schmidt-Catran

IMPRESSUM



DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
http://ideas.repec.org/s/diw/diwwpp.html
http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

# Interactions in Fixed Effects Regression Models

**Marco Giesselmann**[1]

**Alexander Schmidt-Catran**[2]

**Abstract**

An interaction in a fixed effects (FE) regression is usually specified by demeaning the product term. However, this strategy does not yield a genuine within estimator. Instead, an estimator is produced that reflects unit-level differences of interacted variables whose moderators vary within units. This is desirable if the interaction of one unit-specific and one time-dependent variable is specified in FE, but it may yield problematic results if both interacted variables vary within units. Then, as algebraic transformations show, the FE interaction estimator picks up unit-specific effect heterogeneity of both variables. Accordingly, Monte Carlo experiments reveal that it is biased if one of the interacted variables is correlated with an unobserved unit-specific moderator of the other interacted variable. In light of these insights, we propose that a within interaction of two time-dependent variables be estimated by first demeaning each variable and then demeaning the product term. This "double-demeaned" estimator is not subject to bias caused by unobserved effect heterogeneity. It is, however, less efficient than standard FE and only works with T>2.

**Keywords**

panel data, fixed effects, interaction, quadratic terms, polynomials, within estimator

**JEL**

C33; C51

Forthcoming in: Sociological Methods and Research (2020)
https://doi.org/10.1177/0049124120914934

[1] DIW Berlin and Bielefeld University, mgiesselmann@diw.de
[2] Goethe University Frankfurt

## 1. Introduction

Fixed effects (FE) regressions for estimators of variables with intra-unit variation are routinely employed by empirical social scientists, especially when analyzing panel data (see, e.g., the literature surveys in Giesselmann et al. 2015 and in Young and Johnson 2015). The main reason for the popularity of this estimator is its potential to improve causal interpretations (Gangl 2010; Morgan and Winship 2007): because the FE estimator is based solely on variation within units, it automatically controls for all observable and unobservable unit-specific characteristics (Allison 2009; Wooldridge 2010).

Many scholars have dealt with the analytical properties of the FE estimator (e.g., Baltagi 2005; Brüderl and Ludwig 2015; Cameron and Trivedi 2005), with different approaches to its specification in regression frameworks (Andreß et al. 2013; Firebaugh, Warner, and Massoglia 2013; Mandlak 1978), and with its theory-into-practice problems (Halaby 2004; Plümper, Troeger, and Manow 2005). Here, we focus on an issue that has not been dealt with in greater detail, yet; namely the proper specification of interactions in an FE regression. Specifically, we focus on the interaction of two independent variables with intra-unit variation, such as *number of children* and *income* (as modeled by Kühhirt 2012 to explain variations in couples' division of labor). In this article, we refer primarily to the case of FE with longitudinal microdata, although the mechanisms enlightened here also apply to other types of FE models with clustered data.

Basically, an interaction measures how the effect of the independent variable changes with the size of a moderator variable. In ordinary regression frameworks, it is usually tested by including a product term (Allison 1977; Jaccardi and Turrisi 2003). The standard way of specifying such an interaction in an FE regression is to treat the product term as any other variable and, accordingly, to demean it: each realization of the term is subtracted by its unit-specific mean before entering the regression. This strategy is widely used in empirical practice; not only for interactions with one time-constant variable, but also if both interacted variables show variation within units (e.g., Abendroth et al. 2014; Killewald and Gough 2013; Kühhirt 2012; Oesch and Lipps 2013; Schofer and Longhofer 2011). It is also computed by default by statistical programs (Cameron and Trivedi 2009), introduced as desirable specification in methodological discourses (Schunck 2013), and automatically reproduced by an OLS estimation with unit dummies. However, this strategy does not produce a within estimator of the interaction. Hence, it does not control for effect heterogeneity and does not exhibit the desirable statistical properties of FE estimators.[i] It requires the assumption that specified moderators *Z* of a variable *x* are not correlated with time-constant unobserved moderators of *x*.

## 2. Why Demeaning an Interaction Does Not Yield a Within Estimator: Some Intuitive Considerations

The purpose of a within estimator of an interaction between *x* and *z* on *y* is to determine how the within effect of *x* on *y* increases as *z* changes within units. Or, in more technical terms: it shall measure how the intra-unit association between *x* and *y* is related to intra-unit variation in a moderator *z*. The integration of the demeaned interaction term $(z_{it}x_{it} - \overline{(zx)}_i)$ in an FE regression framework yields a coefficient$\hat{\beta}_{dm(zx)}$, which is, however, not a within estimator—it is still identifiable when one of the two variables exhibits only between-unit variance. Thus, the most straightforward "proof" that $\hat{\beta}_{dm(zx)}$ measures between-unit variance is that a unit-specific characteristic *z* (i.e., a variable that is constant within units) can be integrated (and estimated) as moderator of a variable *x*.

Technically, this can be explained by the fact that each realization of a demeaned product term depends on both factors' unit-specific levels: although the term's mean on the unit level is zero, its intra-unit variance depends on the factors' unit-specific means. Specifically, the higher the unit-specific level of one factor, the more strongly the idiosyncrasies of the other are spread in the product term. Thus, the demeaned product describes a distribution of realizations that reflects both the size of the unit-specific components and the idiosyncrasies of the interacted variables.

This algebraic feature of demeaned products allows us to differentiate within-unit effects across categories (or levels) of time-constant characteristics in FE regressions. Therefore, integrating the demeaned product of one time-constant variable and one time-dependent variable into a (FE) regression framework is a standard application of empirical practitioners working with clustered data. However, this feature also defines $\hat{\beta}_{dm(zx)}$ as a between-unit estimator if *both* interacted variables show intra-unit variation.

## 3. Formal Decompositions of FE Interactions

In this section, we support insight on the basis of decompositions of demeaned product terms.

For any measurement {i,t}, the demeaned interaction term $z_{it}x_{it} - \overline{(zx)}_i$ can be written as

$$(1)\quad (z_{it}x_{it}) - \frac{\sum_{t=1}^{T_i} z_{it}x_{it}}{T_i},$$

with every $z_{it}$ and $x_{it}$ consisting of a unit-specific component ($\bar{z}_{i.}$ and $\bar{x}_{i.}$) and a measurement-specific component or, rather, "idiosyncrasy" ($dm(z_{it})$ and $dm(x_{it})$).

Now let us consider three different constellations of variation patterns among the factors *z* and *x*: (1) both factors show no intra-unit variation, (2) only one factor shows intra-unit variation, or (3) both factors show intra-unit variation. How does the (transformed) demeaned interaction term look in these cases and what does this reveal about the properties of its coefficient $\hat{\beta}_{dm(xz)}$?

### 3.1 Case 1: Both Factors *x* and *z* Are Unit-Specific Variables Without Any Intra-Unit Variation

In this case, $x_{it} = \bar{x}_{i.}$ and $z_{it} = \bar{z}_{i.}$ for each measurement {i,t}. Therefore, equation (1) can be written as

$$(2)\ \ (\bar{z}_{i.}\bar{x}_{i.}) - \frac{T_i(\bar{z}_{i.}\bar{x}_{i.})}{T_i} = 0 \text{ for each \{i,t\}.}$$

Thus, demeaning an interaction term with two unit-specific variables eliminates all between-unit variation of both variables, so no moderating influence of *z* on the effect of *x* can be estimated.

### 3.2 Case 2: Only One Factor Exhibits Intra-Unit Variation, While the Other Is Unit-Specific

Let *z* be constant within units and, therefore, $z_{it} = \bar{z}_{i.}$ for all {i,t}. In this case, equation (1) can be written as

$$(3)\ \ (\bar{z}_{i.}x_{it}) - \frac{\sum_{t=1}^{T_i} \bar{z}_{i.}x_{it}}{T_i}.$$

For each measurement {i,t}, $z_{i.}$ is a factor in all added terms in the subtrahend of equation (3). Therefore, it can be factored out, and equation (3) can be written as

$$(4)\ \ (\bar{z}_{i.}x_{it}) - (\bar{z}_{i.}\sum_{t=1}^{T_i} \frac{x_{it}}{T_i})$$

$$(5)\ \ = \bar{z}_{i.}(x_{it} - \sum_{t=1}^{T_i} \frac{x_{it}}{T_i})$$

$$(6)\ \ = \bar{z}_{i.}(x_{it} - \bar{x}_{i.})$$

$$(7)\ \ = \bar{z}_{i.}dm(x_{it}).$$

Thus, if factor *x* of a product term shows intra-unit variation and factor *z* does not, the demeaned term measures between-unit differences of *z* and within-unit differences of *x*. Specifically, equation (7) reveals that $\hat{\beta}_{dm(zx)}$ in this case estimates how the within-unit effect of *x* differs according to between-unit levels of *z*. As noted, this is nothing new and is a characteristic of FE that is widely used in empirical practice. However, comparing cases 1 and 2 reveals a surprising property of demeaned interaction terms: although *z* does not show within-unit variation, the demeaned term contains the unit-specific level of *z* in case 2, whereas it is completely eliminated in case 1. Obviously, in case 2, within-unit variation of *x* activates the unit-specific component of *z* in the demeaned interaction term. This insight has substantial consequences for case 3.

### 3.3 Case 3: Both Factors *x* and *z* Exhibit Intra-Unit Variation

We have seen that two unit-specific variables are completely eliminated in a demeaned interaction term. As soon as one variable shows intra-unit variation, however, this within-unit variation is related to the unit-specific component of the other variable by the demeaning procedure, as shown in equation (7). Therefore, within-unit variation of one factor will cause unit-specific heterogeneity of the time-constant factor to enter into the standard FE estimation of an interaction. Thus, if both variables vary within units, the identification of $\hat{\beta}_{dm(zx)}$ will be based on the moderating effects of the unit-specific levels of both variables.

If both variables are time-dependent, each measurement $x_{it}$ and $z_{it}$ can be regarded as consisting of a unit-specific component ($\bar{z}_{i.}$ and $\bar{x}_{i.}$) and a measurement-specific component ($dm(x_{it})$ and $dm(z_{it})$). Thus, equation (1) can be written as

$$(8)\ \ (\bar{z}_{i.} + dm(z_{it})) * (\bar{x}_{i.} + (dm(x_{it})) - \frac{\sum_{t=1}^{T_i}(\bar{z}_{i.}+dm(z_{it})) * (\bar{x}_{i.}+(dm(x_{it}))}{T_i}.$$

By expanding equation (8), we obtain

$$(9)\ \ \bar{z}_{i.}\bar{x}_{i.} + \bar{z}_{i.}dm(x_{it}) + \bar{x}_{i.}dm(z_{it}) + dm(z_{it})dm(x_{it}) - \frac{\sum_{t=1}^{T_i}\bar{z}_{i.}x_{i.}+\bar{z}_{i.}dm(x_{it})+\bar{x}_{i.}dm(z_{it})+ dm(z_{it})dm(x_{it})}{T_i}.$$

By fractional arithmetic, equation (9) can be transformed into

$$(10)\ \ \bar{z}_{i.}\bar{x}_{i.} + z_{i.}dm(x_{it}) + \bar{x}_{i.}dm(z_{it}) + dm(z_{it})dm(x_{it}) - \frac{\sum_{t=1}^{T_i}\bar{z}_{i.}\bar{x}_{i.}}{T_i} - \frac{\sum_{t=1}^{T_i}\bar{z}_{i.}dm(x_{it})}{T_i} - \frac{\sum_{t=1}^{T_i}\bar{x}_{i.}dm(z_{it})}{T_i} - \frac{\sum_{t=1}^{T_i}dm(z_{it})dm(x_{it})}{T_i}.$$

Factoring out all factors that are constant for each *i*, equation (10) can be written as

$$(11)\ \ \bar{z}_{i.}\bar{x}_{i.} + \bar{z}_{i.}dm(x_{it}) + \bar{x}_{i.}dm(z_{it}) + dm(z_{it})dm(x_{it}) - \bar{z}_{i.}\bar{x}_{i.} - \bar{z}_{i.}\frac{\sum_{t=1}^{T_i}dm(x_{it})}{T_i} - \bar{x}_{i.}\frac{\sum_{t=1}^{T}dm(z_{it})}{T_i} - \frac{\sum_{t=1}^{T}dm(z_{it})dm(x_{it})}{T_i}.$$

As the sum of demeaned values on the unit level is zero, equation (11) can be reduced to

$$(12)\ \ \bar{z}_{i.}dm(x_{it}) + \bar{x}_{i.}dm(z_{it}) + dm(z_{it})dm(x_{it}) - \frac{\sum_{t=1}^{T_i}dm(z_{it})dm(x_{it})}{T_i}.$$

The final transformation in equation (12) reveals that, for each measurement {i,t}, the size of a demeaned interaction term with two variables *z* and *x* showing intra-unit variation depends on the unit-specific levels of both *z* and *x*. Specifically, we see that the size of the demeaned interaction term depends on the products of *(a)* the unit-specific mean in *z* and the demeaned value of *x*; *(b)* the unit-specific mean in *x* and the demeaned value of *z*; and *(c)* both demeaned values of *x* and *z*. In addition, it subtracts the unit-specific mean of *(c)*. Thus, $\hat{\beta}_{dm(zx)}$ measures a combination of *(a)* the moderating influence of level differences in *z* on within variation in *x*, *(b)* the moderating influence of level differences in *x* on within variation in *z*, and *(c)* the moderating influence of within-unit differences in *z* on within-unit variation in *x*. This means that the coefficient of a demeaned interaction term $\hat{\beta}_{dm(zx)}$ does not control for, but includes effect heterogeneity across units.[ii]

## 4. Obtaining the Within Coefficient If Both Variables Vary Within Units

How can we obtain a genuine within estimator of the interaction between two variables *z* and *x* with intra-unit variation? In this case, rather than the demeaned product of two variables, one needs to include a product term of two demeaned variables, as illustrated in equation (13):

(13) $$(z_{it} - \bar{z}_{i.}) * (x_{it} - \bar{x}_{i.}).$$

This term relates intra-unit idiosyncrasies of *x* to intra-unit idiosyncrasies of *z*. Consequently, its coefficient $\hat{\beta}_{dm(z)dm(x)}$ measures how the within-unit association between *x* and *y* changes with intra-unit variation of *z*. This strategy eliminates unit-specific levels of *z* and *x*, thus systematically controlling for effect heterogeneity. However, there is one good reason to object to this conclusion (see, e.g., Schunck 2013), because equation (13) can be expanded and written as

(14) $$z_{it}x_{it} - \bar{z}_{i.}x_{it} - \bar{x}_{i.}z_{it} + \bar{z}_{i.}\bar{x}_{i.}.$$

From this perspective, it does not appear at first glance that unit-specific means actually are eliminated in the product of demeaned values—but they are. This is revealed once we again regard each measurement $x_{it}$ and $z_{it}$ as consisting of a unit-specific component ($\bar{x}_{i.}$ and $\bar{z}_{i.}$) and a measurement-specific component ($dm(x_{it})$ and $dm(z_{it})$). Then, equation (14) can be written as

(15) $$(\bar{z}_{i.} + dm(z_{it}))(\bar{x}_{i.} + dm(x_{it})) - \bar{z}_{i.}(\bar{x}_{i.} + dm(x_{it})) - \bar{x}_{i.}(\bar{z}_{i.} + dm(z_{it})) + \bar{z}_{i.}\bar{x}_{i.}.$$

Using elementary algebra, equation (15) can be written as

(16) $$\bar{z}_{i.}\bar{x}_{i.} + \bar{x}_{i.}dm(z_{it}) + \bar{z}_{i.}dm(x_{it}) + dm(z_{it})dm(x_{it}) - \bar{z}_{i.}\bar{x}_{i.} - \bar{z}_{i.}dm(x_{it}) - \bar{x}_{i.}\bar{z}_{i.} - \bar{x}_{i.}dm(z_{it})) + \bar{z}_{i.}\bar{x}_{i.}$$

(17) $$= dm(z_{it})dm(x_{it}).$$

Transformations in equations (13) to (17) may seem recursive, but they reveal that expanding the product of two demeaned variables leads back to a term that is independent of unit-specific mean values.

Still, there is one substantial problem with $\hat{\beta}_{dm(z)dm(x)}$: the demeaned product of two time-dependent variables does not have a unit-specific mean value of zero (unless the two variables are uncorrelated). Thus, equation (13) still reflects a unit-specific characteristic: because the size of the mean of a product term depends on the degree of covariation among the factors, the unit mean of equation (13) depends on the unit-specific degree of intra-individual correlation among *x* and *z*. Therefore, if an unobserved moderator is related to $Corr_i(x, z)$, its effect will be transported in $\hat{\beta}_{dm(z)dm(x)}$; in order to obtain a within estimator, the product of the demeaned variables must also be demeaned. Then, we arrive at a specification that is functionally equivalent to a specification introduced by Balli and Sørensen (2013) in the context of cross-country analysis:

(18) $$(z_{it} - \bar{z}_{i.}) * (x_{it} - \bar{x}_{i.}) - \frac{\sum_{t=1}^{T_i}(z_{it}-\bar{z}_{i.})*(x_{it}-\bar{x}_{i.})}{T_i}.$$

Alternatively, equation (18) can be written as

(19) $$dm(dm(z_{it})dm(x_{it})).$$

Equations (18) and (19) describe a genuine within transformation: unit-specific mean values of the interacted variables are completely eliminated in the transformed term, and, at the same time, its mean value on the unit level is zero. As a result, heterogeneity in levels, effects, and covariances is completely controlled for in $\hat{\beta}_{dm(dm(z)dm(x))}$, which therefore qualifies as an accurate within estimator of interactions with two time-dependent variables. Unlike the standard FE estimator of an interaction, it does not transport the effect of unobserved correlated effect heterogeneity in the interacted variables. However, compared with standard FE, a substantial part of variance in the moderators is excluded. Thus, the cost of reducing bias may be imprecise estimates. This is also emphasized by the Monte Carlo experiments detailed in the next section.

## 5. Simulations of Different Specifications of FE Interactions

### 5.1 Simulation Setup

We simulated data using the following basic data-generating process (DGP):

(20) $$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 z_{1it} + \beta_3 x_{it} z_{1it} + \beta_4 z_{2it} + \beta_5 x_{it} z_{2it} + u_i + e_{it}$$

where $\beta_0 = 0, \beta_1 = \beta_2 = \beta_4 = 1$ and $\beta_3 = -1$.

$\beta_5$ varies across conditions, as will be explained further below. This data-generating model is a standard unobserved effects model with three variables in its fixed part: *x*, $z_1$, and $z_2$. Each of these variables consists of a unit-specific component ($x_i$, $z_{1i}$, and $z_{2i}$), varying only between units *i*, and a within-unit component varying within units over time ($dm(x_{it})$, $dm(z_{1it})$, and $dm(z_{2it})$).[iii] Thus,

(21) $x_{it} = x_i + dm(x_{it})$
(22) $z_{1it} = z_{1i} + dm(z_{1it})$
(23) $z_{2it} = z_{2i} + dm(z_{2it})$.

All these components—$x_i$, $z_{1i}$, $z_{2i}$, $dm(x_{it})$, $dm(z_{1it})$, and $dm(z_{2it})$—as well as the unit-specific and the idiosyncratic error terms ($u_i$ and $e_{it}$), are drawn from a joint multivariate standard normal distribution and have a mean of *0* and a standard deviation of *1*, except for the error terms, which have a standard deviation of *4*. Both error terms are uncorrelated with each other and with all other variables in the model. The correlations among the variables $x$, $z_1$, and $z_2$ vary across conditions, as explained below.

The DGP contains two interactions with the variable $x$, one with $z_1$, and another one with $z_2$. The variable $z_1$ takes the role of an observed variable, whereas $z_2$ represents an unobserved variable. Thus, the effect of the variable $x$ is moderated by an observed moderator ($z_1$) and an unobserved moderator ($z_2$). Accordingly, the models fitted to the generated data use only the variables $x$ and $z_1$.

What varies between the different DGPs in the basic setting are:

1. **The effects of $\boldsymbol{\beta_5}$.** $\beta_5$ takes the values 0, −0.5, and −1. If $\beta_5$ takes the value 0, there is no unobserved moderation of the effect of *x*. For $\beta_5 = -0.5$ the unobserved moderator effect is half as strong as the observed effect, and for $\beta_5 = -1$ it is of the same strength.
2. **The correlations between the specified and the unobserved unit-specific moderators $z_{1i}$ and $z_{2i}$.** *Corr*($z_{1i}$, $z_{2i}$) takes the values *0*, *0.4*, and *0.8*, representing nonexistent, moderate, and strong correlations, respectively.
3. **The correlations between the specified and unobserved within-unit components $dm(z_{1it})$ and $dm(z_{2it})$.** *Corr*($dm(z_{1it})$, $dm(z_{2it})$) also takes the values *0*, *0.4*, and *0.8*, again representing nonexistent, moderate, and fairly strong correlations, respectively.

For each combination of these conditions we generate data based on the above-mentioned DGP and fit to it a model using the standard FE transformation—as in equation (1), getting the standard FE estimator $\hat{\beta}_{dm(xz)}$—and a model in which the variables are first demeaned then interacted and then demeaned again (as in equation (18), thus obtaining the estimator $\hat{\beta}_{dm(dm(z)dm(x))}$, which we propose to use instead of the standard FE estimator). For each combination of conditions we simulate a population with 1,000,000 observations consisting of 100,000 units, each observed 10 times. From these populations we draw 1,000 samples to which we fit the two different models. Each sample consists of 100 units and 1,000 observations (10 observations per unit).

In an additional simulation, we further vary the number of observations *T* per unit. *T* takes the values *3*, *10*, and *30*. To this additional condition we fit an estimator as in equation (13) ($\hat{\beta}_{dm(z)dm(x)}$), where the two interacting variables are first demeaned and then multiplied, but not demeaned again, and then compare it with the estimator from equation (18) ($\hat{\beta}_{dm(dm(z)dm(x))}$), where the two demeaned variables are interacted and their product is demeaned again. The purpose of these additional simulations is to show why it is necessary to demean the product of two demeaned variables in order to get an unbiased within estimator—that is, why the estimator from equation (18) and not the estimator from equation (13) should be used in practice.

## 5.2 The Proposed vs. the Standard FE Estimator of Interactions

<Figure 1 about here>

Figure 1 shows the results of the simulation study. We present only the estimated interaction effect $\hat{\beta}_3$, as this effect is the sole interest of our paper. The top panel in Figure 1 shows the standard FE estimator and the bottom panel shows the proposed estimator of the demeaned product term of demeaned variables. The correlation between the within-unit components of the specified and the unobserved moderator, $dm(z_{1it})$ and $dm(z_{2it})$, varies over the columns of Figure 1. The x-axes of the single graphs show the correlation between the unit-specific components of the specified and the unobserved moderators, $z_1$ and $z_2$. The y-axes give the estimated effect $\hat{\beta}_3$, where the true effect equals $-1$. Finally, each graph contains three lines representing the different true values of $\beta_5$—that is, the different strengths of the unobserved moderator effect.

As can be seen in the upper panel of Figure 1, the standard FE estimator gives an unbiased estimate of the interaction between $x$ and $z_1$ under two conditions: (a) if there is no interaction between $x$ and the unobserved variable $z_2$ ($\beta_5 = \beta_{xXz2} = 0$), and (b) if the unobserved moderator variable $z_2$ is not correlated with $z_1$, the observed moderator. If there is an unobserved interaction between $x$ and $z_2$ ($\beta_5 = \beta_{xXz2} \neq 0$) and $z_1$ and $z_2$ are correlated, the standard FE estimator is biased, because it carries part of the unobserved effect. In our specific setting there is a negative bias, as the unobserved interaction effect is negative but the correlation between the two moderators $z_1$ and $z_2$ is positive. This bias increases both with the correlation of unit-specific components $Corr(z_{1i}, z_{2i})$ and of within-unit components $Corr(dm(z_{1it}), dm(z_{2it}))$.

In contrast, the estimator proposed in equation (18) (bottom panel of Figure 1) is not biased if the correlation between the observed moderator $z_1$ and the unobserved moderator $z_2$ originates solely from the between-unit components of these variables (left graph). However, if the correlation between $z_1$ and $z_2$ originates from the within-unit components of these variables, the estimator will also be biased.

Nevertheless, the unbiasedness of the proposed estimator under conditions where the standard FE estimator is biased does come at a price: the standard FE estimator is more efficient than the estimator from equation (18).[iv] This is not surprising given that the standard FE estimator uses available between-unit variance in the effect of *x* on *y*, whereas the proposed estimator of demeaned variables does not. Thus, this situation resembles the typical within/between estimator conflict, well known in comparisons of main effects from FE and RE. Likewise, the conclusion is the same: in situations in which time-constant, unobservable moderators are supposed to be correlated with one of the specified moderators *x* or *z*, the standard FE estimator is biased but the proposed estimator is not, so the latter would be the appropriate choice. However, in situations in which the assumption of no correlated, time-constant, unobservable moderators is justified, the standard FE estimator is also unbiased and thus would be the better choice, according to our simulations.[v] Even so, we like to point out that the idea of using a within estimator usually originates from the indication of existing correlated unit-level unobservables. Therefore, an assumption preferring the standard FE estimator of an interaction is somewhat inconsistent with the usual motivation behind using panel data and the respective longitudinal methods (see also Halaby 2004).

### 5.3 The Proposed FE Estimator Without Additional Demeaning of the Term: Why Not Make It Simpler?

Finally, we want to compare the statistical properties of the proposed FE estimator, $\hat{\beta}_{dm(dm(z)dm(x))}$, with an estimator of an undemeaned product term with demeaned variables, $\hat{\beta}_{dm(z)dm(x)}$, as in equation (13). In the simulations outlined in Figure 2, we estimate $\widehat{\beta}_{dm(z)dm(x)}$ and vary the number of observations per unit. To keep it simple, we do not vary the other parameters. Specifically, we consider only the scenario where $Corr(z_{1i},z_{2i})$ and $Corr(dm(z_{1it}), dm(z_{2it}))$ equal *0* and the true effect $\beta_5$ equals −1—that is, a scenario in which both the standard FE and our proposed estimator gave unbiased results.[vi]

<Figure 2 about here>

Obviously, the estimator $\hat{\beta}_{dm(z)dm(x)}$ is biased even if the unobserved moderator $z_{2i}$ is not correlated with the observed moderator $z_{1i}$. The bias diminishes in size with an increasing number of observations per unit. It occurs because the unit-specific mean value of a term obtained by multiplying two demeaned variables is not necessarily zero, even if the idiosyncrasies of the interacted variables are not correlated (and therefore $Corr(dm(x_{1it}), dm(z_{1it})) = 0$). It would be zero if the condition $Corr(dm(x_{1it}), dm(z_{1it})) = 0$ were true for each unit, but with only a few observations per unit (e.g., $T = 3$), it is not even approximately true.

With a growing number of observations per unit (*T = 10, 30*), the data observed within each unit become increasingly representative of the complete data—that is, they reflect $Corr(dm(x_{1it}), dm(z_{1it})) = 0$. In other words, the smaller the number of observations per unit, the more the unit-specific means of the interaction term $dm(x_{1it})*dm(z_{1it})$ deviates from *0* by chance. This adds random variance to the realizations of the interaction term, thus evoking a bias toward *0*, similar to what has been termed *attenuation bias* (also known as *regression dilution*), which occurs because of random measurement errors in the independent variables (Frost and Thompson 2000; Spearman 1904). Note that the proposed estimator in Figure 1 does not show this kind of bias, because an additional demeaning of the product term will eliminate the random variance in unit-level differences.

In addition to the bias toward zero, there may also be a systematic bias in the estimator$\hat{\beta}_{dm(z)dm(x)}$. This can be the case if the product of the two demeaned variables systematically differs from zero (see section 4). Additional simulations, which are not presented here, indicate that such systematic bias occurs once the strength of the intra-individual correlation of *z* and *x* is correlated with an unobserved variable. Compared with the bias toward zero, however, this systematic bias appears to be of minor importance.

In sum, our additional simulation shows that the term $dm(x_{1it})*dm(z_{1it})$ does indeed need to be demeaned in order to obtain a genuine, unbiased within estimator. The estimator $\hat{\beta}_{dm(z)dm(x)}$ is subject to a bias toward zero owing to random variation in its mean. It can also be subject to omitted variable bias if the correlation between $dm(x_{1it})$ and $dm(z_{1it})$ depends on the level of an unobserved effect.

## 6. A Note on FE Interactions with Categorical Variables

After their transformation into (sets of) dummy variables, independent categorical variables are usually treated and interpreted in ways similar to continuously scaled variables in FE regression frameworks (Allison 2009). Therefore, the standard FE estimator of an interaction of categorical variables exhibits the same problems as those outlined for continuous independent variables. Accordingly, genuine within estimators of interactions of dummy variables can be obtained only by demeaning the dummies before they enter the product term.

However, researchers might regard the combinations of two categorical variables as realizations of a new variable—a perspective that is supported by what are referred to in modern statistical software as "factor variables" commands (see, e.g., StataCorp 2015). Instead of multiplicative terms, such commands integrate indicators for each combination of the categories of the variables. In an FE framework, this strategy will naturally result in genuine within estimators of the respective (combined) dummy variables, measuring strictly intra-individual outcome differences relative to the reference category. The fact that these estimators are reproduced by coefficients of standard FE interactions—which are, as argued, *no* genuine within estimators—may seem paradoxical, but it is not:

Generally, differences in within estimators of main effects may originate from interactions of unit-specific characteristics, so relations between main (within) effects of combined indicators allow no inference regarding the within estimator of the underlying interaction. Consequently, although a comparison of FE coefficients of combined indicators reveals differences in a factor's within effects (across categories of the other factor), the differences themselves should not be interpreted as within comparisons. As shown, these differences may be a result of moderating properties of the unit-specific parts of the involved factor.

In other words, intra-individual outcome differences between categories of a combined variable (e.g., the satisfaction gap of a single person between phases of parenthood vs. nonparenthood) may have been produced by moderating features of time-constant characteristics (e.g., by social skills, as a relevant unit-specific determinant of the partnership status). A difference in the satisfaction gap between a single and a nonsingle person therefore is disqualified as a within estimator of an interaction, because it does not strictly relate the within-unit differences in one factor (single vs. nonsingle) to the within-unit changes in the association between the second factor (parent vs. nonparent) and the dependent variable.

## 7. Discussion

By using empirical considerations, formal arguments, and Monte Carlo experiments we have shown that the standard way of specifying an FE interaction of two variables *x* and *z* with intra-unit variation does not always yield an estimator with desirable properties: the coefficient $\hat{\beta}_{dm(zx)}$ is not a genuine within estimator, because it measures a combination of several between-unit and within-unit interactions (see Eq. 12). It will be biased in presence of *correlated effect heterogeneity*, that is if there is an unobserved time-constant moderator of *z* correlated with *x*, or of *x* correlated with *z*. In contrast, by first demeaning the factors and then demeaning the product term ("double demeaning"), unit-specific elements in the products are completely eliminated. Consequently, the estimator $\hat{\beta}_{dm(dm(z)dm(x))}$ is a true within-unit estimator, automatically controls for effect heterogeneity, and therefore yields more consistent results in the presence of correlated unit-specific unobservable moderators. However, its implementation comes at a price—namely, a loss of efficiency compared with standard FE (as illustrated by our Monte Carlo experiments). This is not surprising given that the standard FE estimator uses available between-unit variation in the effects of the independent variables to construct the interaction coefficient (as shown in Eq. 12), whereas the proposed estimator of demeaned variables discards this source of variance. Additionally, it absorbs one additional degree of freedom per unit. Therefore, double demeaning yields imprecise estimates if the interacted variables' within-unit variation is small or the average number of measures on unit level is low. Also, it works only for interactions of time-varying variables and requires more than two measures on the unit level. Such problems are well known from methodological discourses on within-unit estimators (e.g., Allison 2009, p. 23.), but appear to be noticeably severe in the context of interactions; specifically, owing to the omission of all units with T < 3 from the estimation of the interaction.[vii]

If we return to the example presented in the introduction, we see in light of this insight that an FE coefficient of the demeaned product of the variables *number of children* and *income* (explaining, e.g., *couples' division of labour*) is not a within estimator. In addition to intra-individual interdependencies, the standard FE coefficient will also reflect how the within effect of *number of children* on *couples' division of labour* varies among persons with different levels of *income*, and vice versa. Therefore, it transports the moderating influence of income related time-constant unobservables (e.g., class, milieu, cognitive and noncognitive skills, etc.) on the effect of having children on couples' division of labour.

In the empirical literature, these problems of standard FE estimators are rarely addressed. One exception is a study by Schober and Stahl (2016). Focusing on the interaction of two time-dependent variables in a panel regression framework, these authors used the unit-specific mean $\bar{z}_{i.}$ instead of the measurement-specific value $z_{it}$ as moderator. In the example discussed above, this coefficient would tell us how the within effect of having children on division of labour varies across persons at different income levels. In the presence of effect heterogeneity in a panel design, clearly this strategy does not yield a consistent estimator of the time-dependent moderator (it does so only of its unit-specific part). However, as shown in equation (7), it uses only within-unit variance of the unfixed variable, offering a clear and technically accurate interpretation of the coefficient. Thus, although Schober and Stahl (2016) did not specify a within estimator of the interaction, they found a practical solution to the difficulties of testing an interaction in an FE framework. In addition, they offered a technically consistent interpretation and description of their estimator. When we reviewed the major sociological journals, we did not find any other articles that either used a genuine within estimator of interactions or discussed the limitations of the standard FE interaction coefficient accurately. Schofer and Longhofer (2011), for example, measured the effects of policy indicators on the number of associations using a standard FE model. They introduced their estimator as exploiting only "'within-case' variability over time" (p. 559), which is, as shown, not correct for the outlined coefficient of the interaction between *degree of democracy* and *state expansion*. Similarly, Abendroth et al. (2014) asserted that their FE model on determinants of mothers' occupational status "provide stringent tests of within-person change" (p. 10). However, their hypothesis that mothers' age influences the effect of higher-order births was tested by a demeaned interaction term and is therefore *not* based solely on within-unit variation.

Finally, we would like to emphasize that a nonlinear (e.g., quadratic) variable can be regarded as special case of an interaction. Therefore, the mechanisms revealed in this article also apply in such cases: demeaning the product of a quadratic term in an FE framework will cause the influence of the unit-specific levels on the variable's idiosyncrasies to reenter the model,[viii] as shown in equation (12). To obtain a genuine within estimator, the variable must similarly be demeaned before being transformed into its nonlinear form.

**Notes**

---

[i] Basically, the insight that heterogeneity in effects is not automatically controlled for in FE, is not new (e.g. Allison 2009: 29). Here, we reveal the consequences of this omission for interactions.

[ii] This explains algebraically why Balli and Sørensen (2013) find standard FE interactions to be biased in the presence of unobservable unit heterogeneity.

[iii] With regard to the formulas above, the variables $x_i$, $z_{1i}$, and $z_{2i}$ (i.e., the between-unit components) are equivalent to the unit-specific means ($\bar{x}_{i.}, \bar{z}_{i.}$); however, we use a different notation here as they are not derived as means from single observations but are generated directly.

[iv] The standard deviation of the estimated coefficients $\hat{\beta}_{dm(dm(z)dm(x))}$ from the 1,000 samplings is about twice as high for the proposed estimator as for the standard FE estimator. More precisely, on average, over all conditions, the standard deviation is 2.14 times larger for the proposed estimator when compared with the standard FE estimator (min = 1.89, max = 2.47).

[v] In scenarios in which unobserved and observed within components are more strongly correlated than the respective between components, the standard FE estimator even performs better in terms of consistency. However, such scenarios, in which between-unit comparisons provide better counterfactual measures than do within-unit comparisons, are empirically and statistically implausible.

[vi] However, we checked the robustness of findings in additional simulations in which we used different values for the correlation parameters $Corr(z_{1i}, z_{2i})$ and $Corr(dm(z_{1it}), dm(z_{2it}))$.

[vii] We thank our colleague Conrad Ziller (University of Cologne) for leading us to the insight that double demeaning does not work with T=2.

[viii] We are aware that this problem of quadratic terms in FE frameworks was previously revealed in a working paper by McIntosh and Schlenker (2006).

**References**


Abendroth, Anja-Kristin, Matt L. Huffman, and Judith Treas. 2014. “The Parity Penalty in Life Course Perspective: Motherhood and Occupational Status in 13 European Countries.” *American Sociological Review* 79:993-1014.

Allison, Paul D. 1977. “Testing for Interaction in Multiple Regression.” *American Journal of Sociology* 83:144-53.

Allison, Paul D. 2009. *Fixed Effects Regression Models*. Los Angeles: Sage.

Andreß, Hans-Jürgen, Katrin Golsch, and Alexander W. Schmidt. 2013. *Applied Panel Data Analysis for Economic and Social Surveys*. Berlin: Springer.

Balli, Hatice O. and Bent E. Sørensen. 2013. “Interaction Effects in Econometrics.” *Empirical Economics* 45:583-603.

Baltagi, Badi H. 2005. *Econometric Analysis of Panel Data*. Chichester: Wiley.

Brüderl, Josef and Volker Ludwig. 2015. “Fixed-Effects Panel Regression.” Pp. 327-57 in *The Sage Handbook of Regression Analysis and Causal Inference*, edited by H. Best and C. Wolf. Los Angeles: Sage.

Cameron, A. Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cameron, A. Colin and Pravin K. Trivedi. 2009. *Microeconometrics Using Stata*. College Station, TX: Stata Press.

Firebaugh, Glenn, Cody Warner, and Michael Massoglia. 2013. “Fixed Effects, Random Effects, and Hybrid Models for Causal Analysis.” Pp. 113-32 in *Handbook of Causal Analysis for Social Research*, edited by S. L. Morgan. Dordrecht: Springer Science + Business.

Frost, Chris and Simon G. Thompson. 2000. “Correcting for Regression Dilution Bias: Comparison of Methods for a Single Predictor Variable.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 163:173-89.

Gangl, Markus. 2010. “Causal Inference in Sociological Research.” *Annual Review of Sociology* 36:21-47.

Giesselmann, Marco, Carsten Schröder, Johannes Giesecke, John Haisken-DeNew, Anika Rasner, and Jule Specht. 2015. “Editorial: From Panel Data to Longitudinal Analytical Designs: A Note on Contemporary Research Based on Data from the Socio Economic Panel Study (SOEP).” *Schmollers Jahrbuch* 135:1-11.

Halaby, Charles N. 2004. “Panel Models in Sociological Research: Theory into Practice.” *Annual Review of Sociology* 30:507-44.

Jaccard, James and Robert Turrisi. 2003. *Interaction Effects in Multiple Regression* (2nd ed.). Thousand Oaks, CA: Sage.

Killewald, Alexandra and Margaret Gough. 2013. “Does Specialization Explain Marriage Penalties and Premiums?” *American Sociological Review* 78:477-502.

Kühhirt, Michael. 2012. “Childbirth and the Long-Term Division of Labour Within Couples: How Do Substitution, Bargaining Power, and Norms Affect Parents’ Time Allocation in West Germany?” *European Sociological Review* 28:565-82.

Mandlak, Yair. 1978. “On the Pooling of Time Series and Cross Section Data.” *Econometrica* 46:69-85.

McIntosh, Craig T. and Wolfram Schlenker. 2006. “Identifying Non-Linearities In Fixed Effects Models.” San Diego, CA: University of California, San Diego: School of Global Policy & Strategy. Retrieved October 12, 2016 (https://gps.ucsd.edu/_files/faculty/mcintosh/mcintosh_research_identifying.pdf).

Morgan, Stephen L. and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

Oesch, Daniel and Oliver Lipps. 2013. “Does Unemployment Hurt Less If There Is More of It Around? A Panel Analysis of Life Satisfaction in Germany and Switzerland.” *European Sociological Review* 29:955-67.

Plümper, Thomas, Vera E. Troeger, and P. Manow. 2005. “Panel Data Analysis in Comparative Politics: Linking Method to Theory.” *European Journal of Political Research* 44:327-54.

Schober, Pia S. and Juliane F. Stahl. 2016. “Expansion of Full-Day Childcare and Subjective Well-Being of Mothers: Interdependencies with Culture and Resources.” *European Sociological Review* 32: 593-606.

Schofer, Evan and Wesley Longhofer. 2011. “The Structural Sources of Association.” *American Journal of Sociology* 117:539-85.

Schunck, Rainhard. 2013. “Within and Between Estimates in Random-Effects Models: Advantages and Drawbacks of Correlated Random Effects and Hybrid Models.” *The Stata Journal* 13:65-76.

Spearman, Charles. 1904. “The Proof and Measurement of Association Between Two Things.” *The American Journal of Psychology* 15:72-101.

StataCorp. 2015. *Stata 14 Base Reference Manual.* College Station, TX: Stata Press.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). Cambridge, MA: MIT Press.

Young, Rebekah and David R. Johnson. 2015. “Handling Missing Values in Longitudinal Panel Data with Multiple Imputation.” *Journal of Marriage and Family* 77:277-94.

**Figure 1: Standard FE Estimator vs. Proposed Estimator**

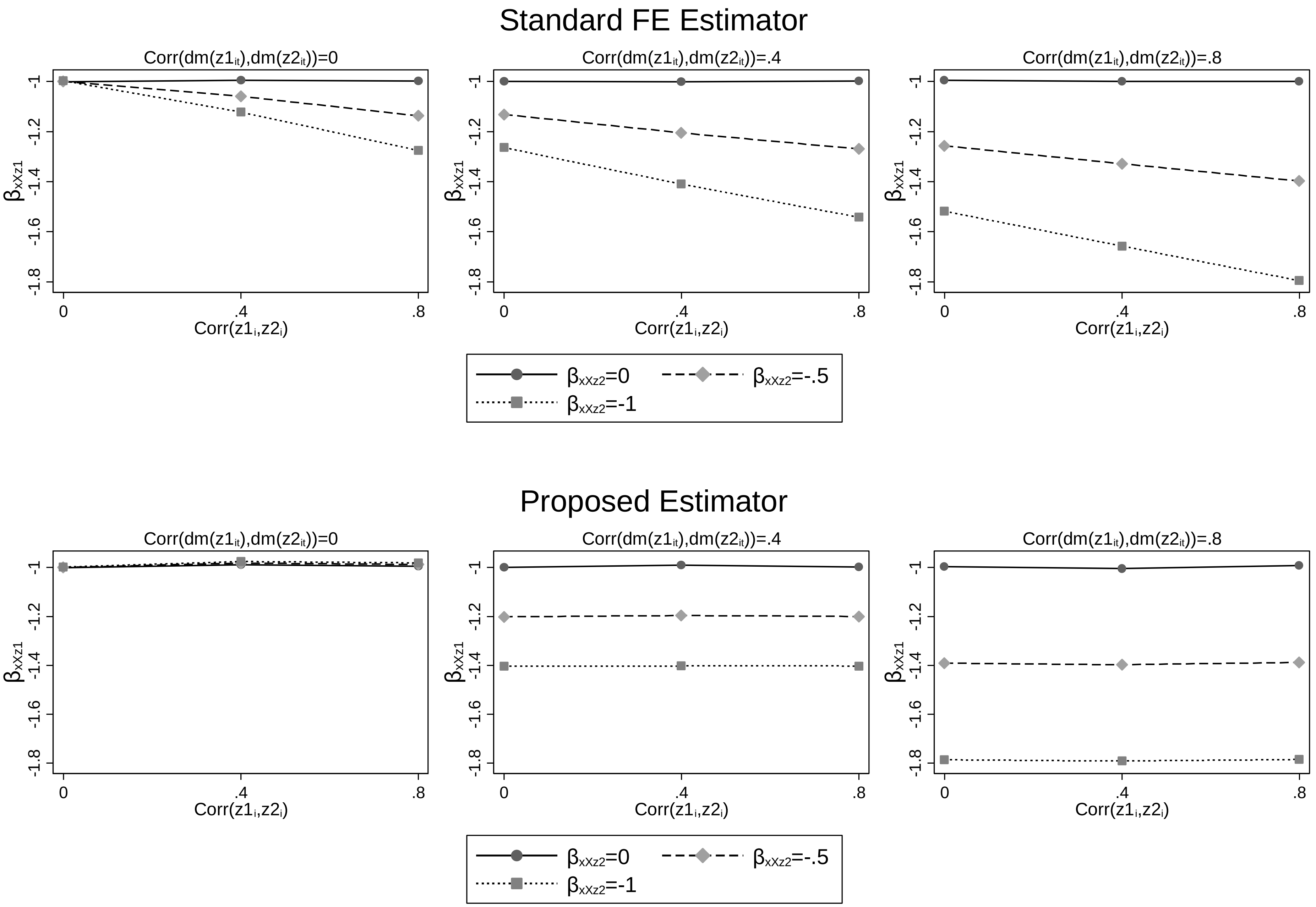

**Figure 2: Effect of varying Ts on estimator without additional demeaning**