Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time Invariant Variables in Fixed Effects Models?

William Greene*
Department of Economics, Stern School of Business, New York University, New York, NY, 10012, USA
wgreene@stern.nyu.edu
http://pages.stern.nyu.edu/~wgreene

February 1, 2010

Abstract

In "Efficient Estimation of Time Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects," Plümper and Troeger (2007), propose a three step procedure for the estimation of fixed effects models that, it is claimed, "provides the most reliable estimates under a wide variety of specifications common to real world data." Their FEVD estimator is startlingly simple, involving three trivial steps, each requiring nothing more than ordinary least squares. Large gains in efficiency are claimed for cases of time invariant and slowly time varying regressors. A subsequent literature has compared the estimator to other estimators of fixed effects models, including Hausman and Taylor's (1981) estimator, also (apparently) with impressive gains in efficiency. The article also claims to provide an efficient estimator for parameters on time invariant variables in the fixed effects model. None of the claims are correct. The FEVD estimator simply reproduces (identically) the linear fixed effects (dummy variable) estimator then substitutes an inappropriate covariance matrix for the correct one. The consistency result follows from the fact that OLS in the FE model is consistent. The "efficiency" gains are illusory. The claim that the estimator provides an estimator for the coefficients on time invariant variables in a fixed effects model is also incorrect. That part of the parameter vector remains unidentified. The "estimator" relies upon turning the fixed effects model into a random effects model, in which case simple GLS estimation of all (now identified) parameters would be efficient among all estimators.

*Helpful discussions of this paper with Neal Beck are gratefully acknowledged. Any remaining errors are, of course, my own.

1. Introduction

The presence of time invariant variables (TIVs) in a panel data regression model poses a vexing problem for the analyst. The usual approach to handling unmeasured heterogeneity in a panel data regression is the fixed effects (FE) model, for which the estimator will be the "least squares dummy variable estimator" (LSDV). The fixed effects approach has some attractive virtues, notably robustness. As is well known, however, it is not possible to include time invariant covariates in a model that is fit by least squares using the individual dummy variables. Using instead simple OLS without accounting for the common effects, "works," but risks serious omitted variable bias if the fixed effects model (with common effects correlated with the regressors) is appropriate, which is usually the case. A random effects (RE) approach, i.e., using GLS instead, allows TIVs but involves an assumption that is rarely palatable, that the common effects are uncorrelated with the regressors. When this assumption fails (as it appears usually to do), the estimator is biased in the same way that OLS estimates are. Plümper and Troeger (2007) (PT), have recently proposed an estimator, labeled FEVD, that appears to solve the longstanding problem of TIVs in an FE model. It is claimed that the procedure greatly improves on the efficiency of LSDV in the fixed effects model, and, along the way solves the problem of non-identification of the coefficients on time invariant variables in this model.

The FEVD estimator is so simple it seems like magic. Like magic, the estimator is illusory. In this note, we will show that the new estimator is algebraically identical to the LSDV estimator so that the claimed efficiency gains cannot be correct. The model and estimator are laid out in Sections 2 and 3. In Section 4, we will prove the equivalence of the FEVD and LSDV estimators and derive the source of the apparent efficiency gains. An example based on a well travelled data set is presented in Section 5 to illustrate the results. The applicable theory of the estimator is developed in Section 6 after the application.

2. The Model

The model is a FE linear regression that contains both time varying and time invariant covariates. Using PT's notation,

$$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + u_i + \varepsilon_{it}\,, \qquad (1)$$

where x_{kit} is a set of K time varying variables, z_{mi} are M time invariant variables and u_i is a set of N-1 unit specific effects. There are N cross section units observed for T periods. The model proposed is a true FE model, so it is assumed, crucially, that $E[u_i|x_{kit},z_{mi}] \neq 0$. It will prove convenient in the discussion that follows to simplify the notation a bit. First, rather than maintain an overall constant and N-1 unit effects, we will formulate the equation with N unit effects and no overall constant – the models are equivalent. Second, we use a convenient matrix formulation. The suggested model becomes

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \qquad [1]$$

where the full NT observations on y_{it} are stacked in $\mathbf{y}$; $\mathbf{X}$ is the full $NT\times K$ matrix on x_{kit}; the N observations on z_{mi} are each repeated T times in each block of the $NT\times M$ matrix $\mathbf{Z}$; and $\mathbf{D}$ is the $NT\times N$ matrix of unit specific dummy variables. (The model and all results to follow are the same if we assume that $\mathbf{D}$ contains a single column of ones and N-1 unit dummy variables.) For convenience, we are assuming a balanced panel – fixed T. The same set of results apply to an unbalanced panel, but at the cost of increased complexity in the notation. In what follows, equation numbers in parentheses refer exactly to the equations in PT while equation numbers in square brackets are used for this paper.

Two distinct cases are suggested by PT. In the case of primary interest here, **Z** consists of a set of TIVs. Any TIV can be written as a linear combination of the *N* dummy variables in **D**. So, for this case, the equation suffers from multicollinearity between **Z** and **D** and γ cannot be estimated apart from α. This is the familiar problem of TIVs in a fixed effects model and is a focus of the paper. In the second case suggested by PT, the columns of **Z** are "slowly changing." But by dint of their changing at all, the variables in **Z** cannot be written as linear combinations of the dummy variables in **D**, which means that the entire set of K+M+N parameters can be estimated consistently and *efficiently* by ordinary least squares, *of necessity including the N dummy variables in the equation*. This case would simply be an FE model with a set of time varying variables, (**X**,**Z**) and the dummy variables, **D**.

The three step procedure and results suggested by PT are intended to apply to both cases. However, in the second case, the claim of increased efficiency for the three step procedure is incorrect. The model is a classical linear regression model with a full rank regressor matrix that is governed by the Gauss-Markov Theorem – the slowly changing variables, **Z**, can be absorbed in **X**. The claimed result in the paper with respect to the slowly changing variables case results from an inappropriate computation of the asymptotic covariance matrix. This will become evident below, where the discussion will encompass both cases. Briefly, the covariance matrix for the three step estimator is computed as if the equation did not contain the dummy variables. This greatly shrinks the elements of the estimated covariance matrix. There is a third possibility suggested by PT. In the slowly changing variables case, their estimator might be a biased estimator with a smaller variance than some competitors, such as Hausman and Taylor (1981). This is indeed a possibility, however in this paper, we are concerned only with the TIV case.

3. The Proposed Estimator

The authors note "This article discusses a remedy to the related problems of estimating time-invariant and rarely changing variables in FE models with unit effects. We suggest an alternative estimator that allows estimating time invariant variables and that is more efficient than the FE model in estimating variables that have very little longitudinal variance. We call this superior alternative "fixed effects vector decomposition (fevd) model."

The proposal consists of three simple steps that involve manipulation of the original data set – no instrumental variables are introduced into the mix, not even the Hausman and Taylor (1981) approach of using the group means of the time varying variables as an additional instrument. It purports to solve the problem of estimating γ while achieving efficiency gains at the same time. In fact, the resulting estimator is algebraically identical to the familiar (original) within groups (dummy variable) estimator. That raises the obvious question of how an identical estimator could become more efficient. Upon closer scrutiny, the efficiency gains claimed in this paper are illusory.

For this proposed estimator, since (we have promised) it can be shown that the estimator is nothing more than ordinary least squares, where do the efficiency gains come from? And, how does an unidentified, inestimable parameter vector become identified and estimable?

The proposal involves the following three step estimation procedure:

Step 1. Estimate α by least squares regression of **y** on **X** and **D**. As they note (p. 5), "We run this FE model with the sole intention to obtain estimates of the unit effects $\hat{u}_i$," which will be $a_i = \hat{\alpha}_i$ in our notation. We proceed from this point using

(4) $$\hat{u}_i = \bar{y}_i - \sum_{k=1}^{K} \beta_k^{FE} \bar{x}_{ki} - \bar{e}_i,$$

"where β_k^{FE} is the pooled-OLS estimate of the demeaned model in equation (3)." By construction, $\bar{e}_i$ = 0, so $\hat{u}_i$ is a_i from the original model. What precisely is contained in a_i depends on the assumptions of the model, as will emerge shortly. Under the strict assumptions in (1), with no further orthogonality assumptions, it must be the case that $\boldsymbol{\gamma} = \mathbf{0}$, and a_i contains α_i plus the sampling error which has mean zero and variance given by (9-18) in Greene (2008). The transition to their (5),

(5) $$\hat{u}_i = \sum_{m=1}^{M} \gamma_m z_{mi} + h_i,$$

requires additional, quite strong assumptions. We will consider this below.

Step 2. Based on (5), the estimated unit effects are regressed on **Z** to obtain an estimator of $\boldsymbol{\gamma}$. The residual h_i is computed from this regression; $h_i = a_i$ - $\mathbf{z}_i'\mathbf{c}^*$, where $\mathbf{c}^*$ is the vector of least squares coefficients in this auxiliary regression. Note, there is a conflict between (5) and this step. The residuals from the regression are not h_i in (5), which are based on the population parameters; the residuals are estimates of h_i based on the least squares "estimates" of $\boldsymbol{\gamma}$.

Step 3. The overall constant, α, coefficient vectors, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ and a new parameter, δ in their

(7) $$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + \delta h_i + \varepsilon_{it}$$

are now estimated by (pooled) ordinary least squares regression of **y** on a constant, **X**, **Z**, and an expanded $NT\times 1$ vector, **h**, in which each h_i is repeated T times.

It is suggested that this three step procedure produces consistent estimators of all of the parameters. Step 3 produces the "correct standard errors." "The third stage also allows researchers to explicitly deal with the dynamics of the time invariant variables." Some simulations demonstrate the superior performance of the estimator. From the conclusions: "Under specific conditions, the vector decomposition model produces more reliable estimates for time invariant and rarely changing variables in panel data with unit effects than any alternative estimator of which we are aware." Once again, our focus at this point is the case of time invariant variables in **Z**.

4 Least Squares Algebraic Results

In spite of the extra layer of interpretation in (5), the regression at Step 3 has the characteristics listed in Table 1, as a result of least squares algebra. That is, the results are not model or data dependent; they will occur exactly as a consequence of the use of least squares. We will prove these results then demonstrate the effect with a familiar data set. The computations are simple and can be replicated with ease with any data set, real or imagined (simulated), and with any modern software.

Table 1. Characteristics of the FEVD Estimator

[a] The overall constant term will be identically zero, in spite of it being attended by an estimated standard error (which is meaningless);

[b] The coefficient estimates on **X** are the original pooled OLS fixed effects coefficient estimates – the same ones obtained at Step 1;

[c] The coefficients on **Z** will be identical to those computed at Step 2;

[d] The coefficient on h_i will identically equal one – as such, its standard error is also meaningless;

[e] The sum of squared residuals and R^2 in the regression at Step 3 are identical to those at Step 1;

[f] The standard errors of the estimates of $\boldsymbol{\beta}$ at Step 3 will appear to be smaller, possibly far smaller, than those computed at Step 1. This is not the result of sampling variability. The matrix computed at this step is systematically too small. We will pursue this result below.

The following employs some basic results for partitioned regression in Greene (2008, pp. 27-29). The estimating equation behind the suggested population model in Step 3 is (7). Our empirical counterpart is

[2] $$\mathbf{y} = \mathbf{Xb} + \mathbf{Zc} + \mathbf{h}d + \mathbf{w}.$$

The (K+M+1) coefficients, ($\mathbf{b}$,$\mathbf{c}$,d) are what will be the least squares (FEVD) solutions, not the population parameters. Thus, $\mathbf{w}$ is the set of least squares residuals, not the population disturbances. (These constructs are mixed at several points in the PT paper, for example in their (4).) The coefficients computed in Step 3 are the OLS solutions in [2]. Because [2] shows the least squares solutions, $\mathbf{X}'\mathbf{w} = \mathbf{0}$, $\mathbf{Z}'\mathbf{w} = \mathbf{0}$ and $\mathbf{h}'\mathbf{w} = 0$, algebraically, not in expectation. First, convert the data in [2] to group mean deviations form by premultiplying by $\mathbf{M_D} = \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$.

[3] $$\mathbf{M_D y} = \mathbf{M_D Xb} + \mathbf{M_D Zc} + \mathbf{M_D h}d + \mathbf{M_D w}$$

On the right hand side, $\mathbf{M_D Z} = \mathbf{0}$ and $\mathbf{M_D h} = \mathbf{0}$ because **Z** and **h** are time invariant so deviations from group means are all zero. That leaves

[4] $$\mathbf{M_D y} = \mathbf{M_D Xb} + \mathbf{M_D w}.$$

The implication is that **b** is the within groups (dummy variables) estimator – [b] in Table 1. We also have that $\mathbf{w} = \mathbf{e}$ from the within groups regression, which proves [e]. We omitted the overall constant in [2], so we have not proved item [a]. To accommodate this, we would add the column of ones to **X**. But, $\mathbf{M_D X}$ would annihilate this column, which would imply (as is obvious in the fixed effects linear regression with a full set of N group dummy variables), the overall constant would be zero.

We now solve for **c** in [2]. Since **b** is determined, the solution will obey

[5] $$\mathbf{y} - \mathbf{Xb} = \mathbf{Zc} + \mathbf{h}d + \mathbf{w}.$$

Premultiply by $\mathbf{Z}'$ to obtain the normal equations and recall that $\mathbf{w} = \mathbf{e}$ from [4]. Then,

[6] $$\mathbf{Z}'(\mathbf{y} - \mathbf{Xb}) = \mathbf{Z}'\mathbf{Zc} + \mathbf{Z}'\mathbf{h}d + \mathbf{Z}'\mathbf{e}.$$

But, $\mathbf{Z'h} = \mathbf{0}$ by construction, and $\mathbf{Z'e} = \mathbf{0}$ because $\mathbf{e}$ is orthogonal to the columns of $\mathbf{D}$ and to every linear combination of the columns of $\mathbf{D}$, including $\mathbf{Z}$. From the within groups regression, $\mathbf{y} - \mathbf{Xb} = \mathbf{a}^* + \mathbf{e}$, where $\mathbf{a}^* = \mathbf{Da}$, so

[7] $$\mathbf{Z'}(\mathbf{a}^* + \mathbf{e}) = \mathbf{Z'Zc}.$$

Since, $\mathbf{Z'e} = \mathbf{0}$, we have

[8] $$\mathbf{Z'a}^* = \mathbf{Z'Zc},$$

which establishes that $\mathbf{c} = (\mathbf{Z'Z})^{-1}\mathbf{Z'a}^* = \mathbf{c}^*$, item [c].

Finally, we now solve for d. Using [2] once again with known $\mathbf{b}$ and $\mathbf{c} = \mathbf{c}^*$, we have

[9] $$\mathbf{y} - \mathbf{Xb} - \mathbf{Zc}^* = \mathbf{h}d + \mathbf{w}.$$

Premultiply by $\mathbf{h'}$, so

[10] $$\mathbf{h'}(\mathbf{y} - \mathbf{Xb} - \mathbf{Zc}^*) = \mathbf{h'h}d + \mathbf{h'w}$$

but $\mathbf{h'w} = 0$ from [2]. As before, $\mathbf{y} - \mathbf{Xb} = \mathbf{a}^* + \mathbf{e}$. Again, $\mathbf{e}$ is orthogonal to every linear combination of the columns of $\mathbf{D}$, including $\mathbf{h}$ (which is time invariant), so

[11] $$\mathbf{h'}(\mathbf{a}^* + \mathbf{e} - \mathbf{Zc}^*) = \mathbf{h'h}d$$

or, since $\mathbf{h} = \mathbf{a}^* - \mathbf{Zc}^*$, $\mathbf{h'}(\mathbf{h} + \mathbf{e}) = \mathbf{h'h} = \mathbf{h'h}d$, or $d = 1$, item [d]. This last result appears in Table 2, in the last row of column (5).

The derivation does not explain the efficiency payoff. How did the standard errors get so small? The appropriate covariance matrix to use for $\mathbf{b}$ is $s^2(\mathbf{X'M_DX})^{-1}$. The result at Step 3 is a submatrix of $fs^2(\mathbf{B'B})^{-1}$ where $\mathbf{B} = (\mathbf{1},\mathbf{X},\mathbf{Z},\mathbf{h})$ and $f = (NT\text{-}K\text{-}N)/(NT\text{-}1\text{-}K\text{-}M\text{-}1)$. The sum of squares from the two regressions are the same, however the variance estimator used by PT appears (incorrectly) to have more degrees of freedom, so f is less than one if M+2 is less than N. In our example, N=595, T=7, K=8 and M=3, so $f = 0.857$. The diagonal elements of this second moment matrix are also smaller than their counterparts in the first one. It can be shown analytically that the second matrix is smaller than the first – the difference is positive definite – fairly easily. We will do so logically instead. In the matrix $\mathbf{X'M_DX}$, the elements are the sums of squares and cross products of the residuals in regressions of the columns of $\mathbf{X}$ on all N of the columns of $\mathbf{D}$. In the submatrix of $\mathbf{B'B}$, the corresponding elements are the sums of squares and cross products of the residuals in the regressions of the columns of $\mathbf{X}$ on only M+2 linear combinations of the columns of $\mathbf{D}$. As long as M+2 is less than N, the sums of squares must be larger – in our example, N is 595 and M+2 is only 5. The sums of squares in these smaller regressions of $\mathbf{X}$ on $(\mathbf{1},\mathbf{Z},\mathbf{h})$ which are only *some* linear combinations of the columns of $\mathbf{D}$, must be larger than their counterparts when $\mathbf{X}$ is regressed on all of the columns of $\mathbf{D}$. When the matrices are inverted, the larger moment matrix becomes the smaller inverted moment matrix. The end result is that $fs^2(\mathbf{B'B})^{-1} << s^2(\mathbf{X'M_DX})^{-1}$ because $f < 1$ and the matrix is systematically smaller. No precise comparison of how much smaller the second matrix is than the first is possible, but the ranking is unambiguous. It is, however, not an appropriate estimator of the asymptotic covariance matrix of $\mathbf{b}_{FEVD} = \mathbf{b}_{FE}$.

That would seem to leave the asymptotic covariance matrix of $\mathbf{c}$, the estimator of the coefficients on $\mathbf{Z}$, to be examined. However, no analysis is possible because γ is not yet an identifiable parameter vector, so no estimator of a covariance matrix for it makes sense. That does not preclude computation of $\mathbf{c}^*$ in Step 2 – it is certainly physically possible to compute the regression. However, there is no meaningful interpretation of the results of this regression *in the context of the fixed effects model*. If the focus is shifted to a random effects model with time invariant variables, then the appropriate comparison would be of this estimate to that obtained by GLS, or some two step method assuming the variance components need to be estimated.

5 A Demonstration

To demonstrate the estimator at work, we will use a simulation based on a "real world" panel data set, that used in the labor market study by Cornwell and Rupert (1988). The data are a balanced panel of observations on 595 individuals for 7 years. The dependent variable of interest in the study is

y = `lwage` = log wage.

The time varying variables are

x = `exp` = experience,
`wks` = weeks worked,
`occ` = a dummy variable for certain types of occupations,
`ind` = a dummy variable for working in industry,
`south, smsa` = dummy variables for living in the south and in an smsa,
`ms` = marital status,
`union` = a union membership dummy.

The time invariant variables are

z = (`fem` = gender, `blk` = race and `ed` = education).

The advantage of pivoting the simulation off a real workd data set is that it is not necessary to make unrealistic (or trivial) assumptions about the interactions among the independent variables. The data on the right hand side of the equation display the characteristics one is likely to encounter in practice.

To produce a simulation with known results but based on a realistic data set, we will proceed as follows: We will use the **X** and **Z** from the observed data. But, we will simulate the dependent variable. By this construction, we will make the data conform exactly to the fixed effects model assumed in the paper, and we will know in advance what the true values of all the parameters in the model are. The specific steps to generate the simulated data are as follows:

(1) Fixed effects linear regression of **y** on **X** and **D**. We retain the predictions from this regression, $yfitFE(i,t)$ and the estimated residual standard deviation, s. The coefficient vector from this regression will be the true coefficients in the model.
(2) Random effects linear regression of **y** on **1**,**X**,**Z**. We retain the coefficients on **Z** from this regression, $\mathbf{c}_{RE}$. Note that the actual values used for these coefficients are not material; we just seek a value that is consistent with the data.
(3) Generate a simulated observation, $ysim(i,t) = yfitFE(i,t) + \mathbf{c}_{RE}'\mathbf{z}_i + s_{FE}\times\varepsilon(i,t)$, where $\varepsilon(i,t)$ is a random draw from the standard normal distribution.

Note that the linear regression of $ysim(i,t)$ on **X**,**D**,**Z** produces exactly the same coefficients and standard errors as the linear regression of $ysim(i,t)$ on **X**,**D** because, as noted, **Z** is a linear combination of the columns of **D**, so least squares estimates its coefficients as zeros.

Thus, *ysim*(i,t) satisfies exactly the assumptions of the model; it is generated by a true fixed effects model with time invariant variables that actually have nonzero coefficients. The disturbances are true random noise, homoscedastic and uncorrelated across observations. The nonzero coefficients on **Z** cannot be estimated in the presence of the dummy variables, but they are embedded in the data nonetheless. The true values of the coefficients used to simulate the data are shown in Table 2 in parentheses under the estimated parameters. Note that the correct values for the standard errors of the fixed effects estimator are also known. Since the disturbances were simulated from a normal distribution with a known standard deviation, s_{FE}, the actual, correct covariance matrix for the fixed effects estimator (conditioned on **X**) is $s_{FE}^2(\mathbf{X}'\mathbf{M_D}\mathbf{X})^{-1}$. These true standard errors are also shown in parentheses in Table 2. Computer code for simulating the data and computing the estimates is given in the Appendix.

The results that appear in Table 2 are to be expected – the equality of the coefficients at Step 3 to those in Steps 1 and 2 was shown algebraically. The payoff is the comparison of the standard errors in the third regression compared to those in the first regression, that is, column (6) vs. column (2), and in the nonzero coefficients on **Z** in column 5. The standard errors have fallen substantially, by factors ranging as high as 6. The population values of γ are shown in column (1). As noted, these are not estimable. The coefficients in column (3) that arguably should be estimates of them are quite far off. However, any resemblance would be coincidental.

The evidence of items [a]-[e] in the results in Table 2 is not a contrivance nor is it a peculiarity of these data. Like results will reappear in any panel data set that is manipulated likewise. We have encountered numerous applications of this method in the recently received literature, including, Akhter and Daly (2009), Alemán (2008), Brück and Peters (2009), Buckley and Schneider (2007), Caporale et al. (2009), Davis (2009), Hansen (2009), Mainwaring and Pérez-Liñán (2008), Sova et al. (2009) and Worrall (2008). The striking reappearance of $\mathbf{b}_{FE}$ in tables of results that present $\mathbf{b}_{FEVD}$ seems not to attract any attention. Likewise, the simple recreation of **c*** as **c** in the second and third step regressions seems unremarkable. Attention in the studies we have seen is focused on the standard errors such as shown in column (6) of our table, which are unambiguously too small, regardless of the data set in use.

6. The Actual Model and Estimators of Its Parameters

The preceding established the equality of $\mathbf{b}_{FRVD}$ and $\mathbf{b}_{LDDV}$ and that the second and third step "estimators" of $\boldsymbol{\gamma}$ are identical. We also established algebraically that the asymptotic covariance matrix computed for the estimator of $\boldsymbol{\beta}$ at Step 3

$$\text{Est.Var}[\mathbf{b}_{FEVD}] = s_{FEVD}^2\,(\mathbf{X}'\mathbf{M}_{\mathbf{1,Z,h}}\mathbf{X})^{-1}$$

must be smaller – every diagonal element is smaller – than the covariance matrix computed at Step 1,

$$\text{Est. Var}[\mathbf{b}_{LSDV}] = s_{LSDV}^2\,(\mathbf{X}'\mathbf{M_D}\mathbf{X})^{-1}.$$

The scale factor, s_{FEVD}^2 is smaller than s_{LSDV}^2 and the former matrix is unambiguously smaller than the latter. The algebraic result is shown after [11]. Two issues remain to settle. First, the much simpler of the two, the preceding results have not established that the estimator of the variance of the FEVD estimator is inappropriate; we have only established that it smaller than the one computed for the LSDV estimator. Second, the appearance of an estimator of $\boldsymbol{\gamma}$ in a model in which, by construction, it should be unidentified is bemusing. We consider both of these in turn.

Table 2. FEVD Three Step Estimation. (Population values in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)
	Step 1 OLS Fixed Effects		**Step 2 OLS, No Constant**		**Step 3 OLS, 1:X,Z,h**	
LHS Var.	LWAGESIM		a_i		LWAGESIM	
R^2	0.95976		n/a*		0.95976	
e′e	79.37106		753.80535		79.37106	
s	0.14927 (0.153221)		1.12841		0.13826 (0.153221)	
Deg.Fr.	3562		592		4152	
	Estimate	Std.Error	Estimate	Std.Error	Estimate	Std.Error
Constant					0.00000	0.04508 (0.00000)
EXP	0.09517 (0.09658)	0.00116 (0.00119)			0.09517	0.00059 (0.00119)
WKS	0.00081 (0.01114)	0.00059 (0.00060)			0.00081	0.00043 (0.00060)
OCC	-0.02384 (-0.02486)	0.01353 (0.01389)			-0.02384	0.00585 (0.01389)
IND	0.02358 (0.02076)	0.01517 (0.01557)			0.02538	0.00467 (0.01557)
SOUTH	-0.00572 (-0.00320)	0.03368 (0.03458)			-0.00572	0.00497 (0.03458)
SMSA	-0.01286 (-0.04373)	0.01908 (0.01958)			-0.01286	0.00485 (0.01958)
MS	-0.05438 (-0.03026)	0.01864 (0.01914)			-0.05438	0.00815 (0.01914)
UNION	0.01547 (0.03416)	0.01465 (0.01504)			0.01547	0.00509 (0.01504)
FEM	(-0.30293)		-0.38338	0.15101	-0.38338	0.01019
ED	(0.10966)		0.47175	0.00381	0.47175	0.00224
BLK	(-0.22565)		-0.09035	0.18337	-0.09305	0.00891
H	(0.00000)				1.00000	0.00650

*The regression does not contain a constant term, so R^2 is not computed.

For the first result, from (1),

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},\ E[\boldsymbol{\varepsilon}|\mathbf{X},\mathbf{Z}] = \mathbf{0},\ E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X},\mathbf{Z}] = \sigma^2\mathbf{I}.$$

The LSDV estimator is

$$\mathbf{b}_{\text{LSDV}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{M}_\mathbf{D}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_\mathbf{D}\boldsymbol{\varepsilon}.$$

It is immaterial whether $\boldsymbol{\alpha}$ is correlated with **Z** and **X** or not (i.e., whether the model is an FE or an RE model). The textbook result is that the correct covariance matrix is given by

$$\text{Var}[\mathbf{b}_{\text{LSDV}}|\mathbf{X},\mathbf{Z}] = \sigma^2(\mathbf{X}'\mathbf{M}_\mathbf{D}\mathbf{X})^{-1}.$$

As shown earlier, the FEVD estimator is using $f\times s_{\text{LSDV}}^2$ to estimate σ^2 where $f = (NT - N - K)/(NT - K - M - 2) < 1$. As N increases, f converges to $(T\text{-}1)/T$. That is, the downward bias in the estimator of σ^2 does not go away, and is worse the shorter is the panel. As shown earlier, the matrix used for computing the variance of the FEVD estimator is also systematically smaller, and the downward bias does not vanish as N increases. The end result is that the estimated covariance matrix for the FEVD estimator of $\boldsymbol{\beta}$ is always too small. By how much is data and application specific.

The authors propose the “estimator” at Step 3 as a method of estimating the parameters $\boldsymbol{\gamma}$ in a fixed effects model that contains TIVs, i.e., their equation (1). The point that seems to be overlooked in the substantial literature that this proposed estimator has inspired is that in the fixed effects model, if $\boldsymbol{\alpha}$ is assumed to exist *as the set of fixed effects*, then γ *does not exist*, so it cannot be estimated, efficiently or otherwise. Reconsider the original model,

$$[1] \qquad \mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \mathbf{D}\alpha + \varepsilon.$$

As noted earlier, **Z** is a linear combination of the columns of **D**, which means that **Z** may be written as **DA** for some $N \times M$ matrix **A** with full column rank $M < N$. Thus, the regression model is

$$[1'] \qquad \begin{aligned} \mathbf{y} &= \mathbf{X}\beta + \mathbf{DA}\gamma + \mathbf{D}\alpha + \varepsilon \\ &= \mathbf{X}\beta + \mathbf{D}(\mathbf{A}\gamma + \alpha) + \varepsilon \\ &= \mathbf{X}\beta + \mathbf{D}\alpha^A + \varepsilon \end{aligned}$$

for some $\boldsymbol{\alpha}^A$. The well known implication is that it is not possible to estimate γ and $\boldsymbol{\alpha}$ separately. Only the preceding linear mixture of the two is estimable. This is a pure example of multicollinearity. It is logically identical to the regression "model," $\mathbf{y} = \mathbf{x}_1\beta + \mathbf{x}_2\gamma + \boldsymbol{\varepsilon}$, in which $\mathbf{x}_2 = 2\mathbf{x}_1$. In such a case, even if the model were "correct," it is not possible to fit it by least squares. One must either assume that $\beta = 0$ or $\gamma = 0$ (or some other known fixed value) or that the simple regression of **y** on $\mathbf{x}_1$ estimates $(\beta + 2\gamma)$. No other construction is possible. Returning to our [1], the solution always employed is to assume $\gamma = \mathbf{0}$, and drop **Z** from the model.

The unconvinced reader will now point to PT's

$$(6) \qquad \alpha = \mathbf{Z}\gamma + \mathbf{h}$$

to argue the opposite. The problem is that (6), like (7), is incorrect. The vector of dummy variable coefficients *in the fixed effects model* is not equal to $\mathbf{Z}\gamma$ plus a disturbance that is uncorrelated with **Z**. That is the point of the model. It is not even the case if it is assumed that α_i is uncorrelated with $\mathbf{x}_{it}$. It will be the case if it is assumed that α_i is uncorrelated with **Z**. But, this is not an assumption in the fixed effects model – the crucial assumption of the FE model is that the common effects can be correlated with the regressors, all of them, TIV or not.

We can obtain a counterpart to (6) if it is assumed at the outset that

$$[12] \qquad \begin{aligned} y_{it} &= \alpha + \mathbf{x}_{it}'\beta + \alpha_i + \varepsilon_{it}, \\ \alpha_i &= \mathbf{z}_i'\gamma + \eta_i, \end{aligned}$$

where η_i is *uncorrelated* with both $\mathbf{x}_{it}$ and $\mathbf{z}_i$. But, this is a random effects model with time invariant variables in it, not a fixed effects model. The difference is crucial. This is not a matter of using OLS vs. LSDV, or any other particular estimator. It is an assumption of the model. The reduced form is

$$y_{it} = \alpha + \mathbf{x}_{it}'\beta + \mathbf{z}_i'\gamma + u_i + \varepsilon_{it},$$

where ε_{it} is as before and now u_i is a random effect. This model is estimable, consistently albeit inefficiently by OLS, and efficiently by GLS or feasible, two step FGLS. When the model is stated with these assumptions, then the three step estimator proposed by Plumper and Troeger does, indeed, estimate $\boldsymbol{\beta}$ and γ. But, that has only been made possible by the additional assumption that the common effects are uncorrelated with the time invariant variables, an assumption that is not part of the fixed effects specification.

The proposed estimator of $\boldsymbol{\gamma}$ is enabled by assuming that the model is a hybrid of the fixed and random effects models. The identifying restriction is that η_i in [12] is uncorrelated with $\mathbf{z}_i$. It is not necessary to assume that η_i is uncorrelated with $\mathbf{x}_{it}$. Thus, the PT model resembles the specification of Hausman and Taylor (1981) where it is assumed (equivalently) that the common effect is uncorrelated with *some* of the variables in $\mathbf{x}_{it}$ and *some* of the variables in $\mathbf{z}_i$. In the PT model, the counterpart is that the effect is uncorrelated with *all* of the variables in $\mathbf{z}_i$ - we use M orthogonality conditions to identify the M parameters in $\boldsymbol{\gamma}$ - and *none* of the variables in $\mathbf{x}_{it}$. To pursue our earlier metaphor, this subtle assumption is how the rabbit gets into the magician's hat. This is the device that identifies the otherwise unidentified $\boldsymbol{\gamma}$. Plümper and Troeger (2007, page 6) make reference to this result where they state " By design, h_i is no longer correlated with the vector of z variables. *If the time invariant variables are assumed to be orthogonal to the unobserved unit effects – i.e., if the assumption underlying our estimator is correct* – the estimator is consistent. If this assumption is violated, the estimated coefficients for the time invariant variables are biased..." (Emphasis added.) This is, in fact, the crucial assumption, but it is not made at any point before this statement. (The discussion also mixes h_i and η_i – in their construction, their h_i is orthogonal to $\mathbf{Z}$ by construction as a least squares residual, but this does not establish the orthogonality of the true unit effects from $\mathbf{Z}$.) However, the central results of this paper hold regardless of this assumption: (1) the FEVD estimator is just LSDV and (2) there is no efficiency gain over LSDV regardless of whether this assumption is met or not.

The proposed estimator of $\boldsymbol{\gamma}$ is that in Step 3, using ordinary least squares. The estimator of the asymptotic covariance matrix based on Step 3 is

$$\mathbf{V}_{(3)} = \frac{\mathbf{e}'\mathbf{e}}{NT - K - M - 2}(\mathbf{Z}'\mathbf{M}_{\mathbf{1},\mathbf{X},\mathbf{h}}\mathbf{Z})^{-1}$$

where

$$\mathbf{M}_{\mathbf{1},\mathbf{X},\mathbf{h}} = \mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}',\ \mathbf{G} = (\mathbf{1},\mathbf{X},\mathbf{h}).$$

For our example, these are the standard errors shown at the bottom of column (6) in Table 2. However, the estimator of $\boldsymbol{\gamma}$ at Step 3 is numerically (algebraically) identical to the estimator computed at Step 2, once again using OLS. Based on this regression, the estimated asymptotic covariance matrix for the Step 2 estimator would be

$$\mathbf{V}_{(2)} = \frac{\mathbf{h}'\mathbf{h}}{N - M - 1}(\mathbf{Z}'\mathbf{Z})^{-1}.$$

These would be the standard errors in column (4). No obvious comparison of these two covariance matrices is possible. The matrix part in $\mathbf{V}_{(2)}$ is unambiguously smaller than that in $\mathbf{V}_{(3)}$. However, the scale factor could go either way. Note in Table 2, the estimated standard errors in column (4) are considerably larger than their counterparts in column (6). But, the comparison is a moot point. Under the assumptions of the model in [12], neither of these matrices is appropriate.

The fixed effect estimator of α_i is given by the result in (4), where $\overline{e}_i = 0$ for every i and $\boldsymbol{\beta}^{FE}$ is actually $\mathbf{b}$. Regardless of whether one views the model as the FEM in (1) or the REM in [12], a_i is not a function of $\mathbf{Z}$; $\mathbf{Z}$ has been swept out by taking deviations from means. The estimator of α_i is

$$\begin{aligned} a_i &= \alpha_i + \text{sampling error} \\ &= \alpha_i + v_i \end{aligned}$$

where the expected value of v_i is 0 – since a_i is unbiased. If we now base our interpretation of the model on [12], then

$$a_i = \mathbf{z}_i'\gamma + \eta_i + v_i$$

The variance of a_i around its mean (which would be $\mathbf{z}_i'\gamma+\eta_i$) is given in Greene (2008), (9-18);

$$[13] \qquad Var[v_i \mid \mathbf{X}] = \frac{\sigma_\varepsilon^2}{T} + \bar{\mathbf{x}}_i'\left[\sigma_\varepsilon^2(\mathbf{X}'\mathbf{M_D}\mathbf{X})^{-1}\right]\bar{\mathbf{x}}_i\,.$$

Combining terms, then, once again, *under the model assumptions*,

$$[14] \qquad \mathrm{Var}[a_i|\mathbf{X},\mathbf{Z}] = \sigma_\eta^2 + \sigma_\varepsilon^2\left\{\frac{1}{T} + \bar{\mathbf{x}}_i'\left[(\mathbf{X}'\mathbf{M_D}\mathbf{X})^{-1}\right]\bar{\mathbf{x}}_i\right\}$$

The regression implied by PT's reformulation of the model is heteroscedastic. The appropriate asymptotic covariance matrix would be

$$\text{Asy.Var}[\mathbf{c}^*|\mathbf{X},\mathbf{Z}] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1},$$

based on the second step, where $\boldsymbol{\Omega}$ is a diagonal matrix containing the elements in [14]. The matrix computed at Step 3 is irrelevant, and bears no resemblance to this or, necessarily, to $\mathbf{V}_{(2)}$. The standard errors in Step 3 are inappropriate. Those in Column (4) of Table 2 are also, but they may resemble the correct result. Since the different variances do not actually contain $\mathbf{Z}$, the computation assuming homoscedasticity with respect to $\mathbf{Z}$ may not be too far off. To investigate for our example, we computed a White, heteroscedasticity corrected, robust covariance matrix for the regression in Step 2. The estimated standard errors are (0.14308,0.00374,0.20643), which are quite close to the naive estimates of (0.15101,0.00381,0.18337) reported in Table 2. They are, however, far larger than those reported in column (6) of Table 2.

The conclusion to this discussion is that the claimed precision of the estimator of γ based on Step 3 is incorrect. The comparison is based on the wrong matrix; it should be based on Step 2, not Step 3.

7. Conclusions

The FEVD estimator proposed by Plümper and Troeger (2007) is illusory. The development of the estimator exploits an interesting algebraic result that reaches an old conclusion via a new path – the estimator is the original least squares dummy variable estimator. The claimed efficiency gains under their assumptions are produced by using an erroneous result, equation (7), to motivate an incorrect covariance matrix, both for estimation of β and for γ. The existence of the estimator for γ hangs on a crucial orthogonality assumption that the analyst may or may not be comfortable with. Assuming they are, then FEVD is a consistent estimator, but the researcher needs to be careful that the covariance matrix that seems to be appropriate (at Step 3) is unambiguously too small. There is a simple remedy for this suggested in the preceding – namely using only Steps 1 and 2 and not computing Step 3 at all. This conclusion is based on the assumptions of the model. For more general cases in which the orthogonality conditions are not met, we must analyze FEVD as an inconsistent estimator with a possibly smaller variance than some competitors such as Hausman and Taylor (1981). However, regardless of this extension, the result remains that Step 3 takes an existing estimator and produces an incorrect covariance matrix – Step 3 should not be carried out regardless of the model assumptions. The full set of results for FEVD are obtained at Steps 1 and 2. The LSDV estimator of β and the asymptotic covariance matris are correctly estimated at Step 1. The estimator of γ coupled with the White robust covariance matrix obtained at Step 2 are appropriate if the orthogonality assumption is met, and are meaningless if not.

References

Akhter, S. and K. Daly, "Finance and Poverty: Evidence from Fixed Effect Vector Decomposition," *Emerging Markets Review*, 10, 2009, pp. 191-206

Alemán, J., " Labor Market Deregulation and Industrial Conflict in New Democracies: A Cross-National Analysis, *Political Studies*, 56, 4, 2008, pp. 830-856.

Brück, T. and H. Peters, "20 Years of German Unification: Evidence on Income Convergence and Heterogeneity," IZA Discussion Paper 4454, 2009.

Buckley, J. and M., Schneider, *Charter Schools: Hope or Hype?,* Princeton University Press, Princeton, 2007.

Caporale, G., C. Rault, R. Sova, A. Sova, "On the Bilateral Trade Effects of Free Trade Agreements Between the EU-15 and the CEEC-4 Countries, *Review of World Economics*, 145, 2, 2009, pp. 189-206.

Cornwell, C. and P. Rupert, "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variable Estimators," *Journal of Applied Econometrics*, 3, 1988, pp. 149-155.

Davis, D., "Price and Promotion Effects of Supermarket Mergers," Working Paper, Department of Economics, South Dakota State, University, 2009.

Greene, W., *Econometric Analysis*, Prentice Hall, 2008

Hansen, M., "Using Private Contracts to Create Adoptions from Foster Care, Working Paper No. 2007-03, Department of Economics, American University, 2007.

Hausman, J. and W. Taylor, "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1981, pp. 1377-1398.

Mainwaring, S. and A. Pérez-Liñán, "Regime Legacies and Democratization: Explaining Variance in the level of Democracy in Latin America, 1978—2004," Political Science Working Paper # 354, 2008, Kellogg School, Notre Dame.

Plümper, T. and V. Troeger, "Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects," *Political Analysis*, 15, 2, 2007, pp. 124-139.

Sova, R., L. Albu, I. Stancu, A. Sova, Patterns of Foreign Direct Investment in the New EU Countries," *Romanian Journal of Economic Forecasting* – 2/2009, pp. 42-51.

Worrall, J., "Racial composition, unemployment, and crime: Dealing with inconsistencies in panel designs," *Social Science Research*, 37, 3, 2008, pp. 787-800

Appendix: NLOGIT Simulation and Estimation Commands

Commands for carrying out the computations are as follows:

```
? Read in Cornwell and Rupert panel data set. Then generate simulated data.
? X = time varying variables, Z = Time invariant variables
  namelist ; x = exp,wks,occ,ind,south,smsa,ms,union $
  namelist ; z = fem,ed,blk $
? Obtain the predictions from the FEM. True coefficients by LSDV estimator.
? Variable LWF is the prediction from this estimated equation
  regress  ; lhs = lwage ; rhs = x ; panel ; pds=7 ; fixed effects ; keep = lwf $
  matrix   ; btruefe = b $  True coefficients
  calc     ; list ;  struefe = s $  Display and catch true sigma for disturbances.
? Obtain a set of coefficients for the TIVs from an REM.
  regress  ; lhs = lwage ; rhs = z,one,x ; panel ; pds=7 ; random effects $
  matrix   ; btiv = b(1:3) $
? Simulated data are obtained by adding disturbances to prediction.
? Add in an effect for the TIVs using the true coefficients.
  calc     ; ran(1234567) $  Set seed for RNG for replicability
  create   ; lwagesim = lwf + btiv(1)*fem + btiv(2)*ed + btiv(3)*blk + struefe*rnn(0,1)$
? True asymptotic covariance matrix differs only by s-squared. Show results
  regress  ; lhs=lwagesim;rhs=one,x;panel;pds=7; fixed effects $
  matrix   ; truevc = {struefe^2/ssqrd}*varb $
  matrix   ; stat(btruefe,truevc)$  These are the theoretically correct values.
? Now compute FEVD estimates using simulated data
  regress  ; lhs=lwagesim;rhs=x;panel;pds=7;fixed effects $ (Step 1)
  create   ; ai=alphafe(_stratum)$ (To stretch the a vector to NT length)
  regress  ; lhs=ai;rhs=z;res=hi $ (Step 2 computes hi for Step 3)
  regress  ; lhs=lwagesim;rhs=one,x,z,hi $  (Step 3 OLS regression)
  reject   ; year > 1 $  (Redo step 2 for right s.e.s using only N obs.)
  regress  ; lhs = ai ; rhs = z $  (Naive estimator of covariance matrix)
  regress  ; lhs = ai ; rhs = z ; hetero $ (Use White estimator instead)
```

The results can be reproduced with any contemporary software; they require only linear least squares regressions. Some small differences will occur across implementations because we used simulated data and random number generators differ across programs. Results that will be identical across packages can be obtained by skipping the simulation and using the original data. This is done in the preceding code by proceeding directly to the computation of the FEVD estimators and using `lwage` rather than `lwagesim` in the two regressions where it appears.