Political Analysis (2011) 19:123–134
doi:10.1093/pan/mpq026

On the Fixed-Effects Vector Decomposition

Trevor Breusch, Michael B. Ward, Hoa Thi Minh Nguyen, and Tom Kompas
Crawford School of Economics and Government, The Australian National University, Canberra, ACT 0200, Australia
e-mail: trevor.breusch@anu.edu.au (corresponding author), michael.ward@anu.edu.au, hoa.nguyen@anu.edu.au, tom.kompas@anu.edu.au

This paper analyzes the properties of the fixed-effects vector decomposition estimator, an emerging and popular technique for estimating time-invariant variables in panel data models with group effects. This estimator was initially motivated on heuristic grounds, and advocated on the strength of favorable Monte Carlo results, but with no formal analysis. We show that the three-stage procedure of this decomposition is equivalent to a standard instrumental variables approach, for a specific set of instruments. The instrumental variables representation facilitates the present formal analysis that finds: (1) The estimator reproduces exactly classical fixed-effects estimates for time-varying variables. (2) The standard errors recommended for this estimator are too small for both time-varying and time-invariant variables. (3) The estimator is inconsistent when the time-invariant variables are endogenous. (4) The reported sampling properties in the original Monte Carlo evidence do not account for presence of a group effect. (5) The decomposition estimator has higher risk than existing shrinkage approaches, unless the endogeneity problem is known to be small or no relevant instruments exist.

1 Introduction

We analyze the properties of a recently introduced methodology for panel data, known as fixed-effects vector decomposition (FEVD), which Plümper and Troeger (2007a) developed to produce improved estimates in cases where traditional panel data techniques have difficulty. Researchers in many fields seek to exploit the advantages of such panel data. Having repeated observations across time for each group in a panel allows one, under suitable assumptions, to control for unobserved heterogeneity across the groups that might otherwise bias the estimates. Mundlak (1978) demonstrated that a generalized least squares approach to unobserved group effects, which treats them as random and potentially correlated with the regressor, gives rise to the traditional fixed-effects (FE) estimator. However, FE is a blunt instrument for controlling for correlation between observed and unobserved characteristics because it ignores any systematic average differences between groups. Thus, any potential explanatory factors that are constant longitudinally (time-invariant) will be ignored by the FE estimator. Likewise, any explanatory variables that have little within variation (i.e., slowly changing over the longitudinal dimension) will have little explanatory power and will result in imprecise coefficient estimates that have large standard errors (SEs).

Hausman and Taylor (1981) had previously shown that a better estimator than FE is available if some of the explanatory variables are known to be uncorrelated with the unobserved group effect, thus described as exogenous explanatory variables. The Hausman-Taylor (HT) estimator is an instrumental variables (IV) procedure that combines aspects of both FE and random-effects estimation. Given a sufficient number of exogenous regressors, the HT procedure allows time-invariant variables to be kept in the model. It also provides more efficient estimates than FE for the coefficients of the exogenous time-varying variables. The downside of the HT estimator resides in specifying the exogeneity status for each of the time-varying and time-invariant variables in the model. In many practical applications such detailed specification is onerous.

Plümper and Troeger introduced FEVD as an alternative that seemed to be superior to HT because it requires fewer explicit assumptions yet seemed to always have more desirable sampling properties. Like the FE estimator, and unlike HT, FEVD does not require specifying the exogeneity status of the explanatory variables. Like the HT estimator, and unlike FE, the FEVD procedure gives coefficient estimates for

Authors' note: Supplementary materials for this article are available on the *Political Analysis* Web site.

time-invariant (and slowly-changing) variables as well as the time-varying variables. Plümper and Troeger motivated the FEVD procedure on heuristic grounds, and advocated it on the strength of favorable results in a Monte Carlo simulation study. In particular, the simulation indicated that FEVD has superior sampling properties for time-invariant explanatory variables.

Although the FEVD procedure comes out of the empirical political science literature, it is rapidly finding application in many other areas including social research and economics. At last count, there were well over 200 references in Google Scholar to this emerging estimation methodology. Several empirical studies report SEs for FEVD-based estimates that are strikingly smaller than estimates based on traditional methods. There is, however, little formal analysis of the FEVD procedure in this literature.

The present paper is a remedy to the lack of formal analysis. We demonstrate that the FEVD coefficient estimator can be equivalently written as an IV estimator, which serves to demystify the nature of the three-stage FEVD procedure and its relationship with other estimators. As one immediate benefit, the IV representation allows us to draw on a standard toolkit of results.

First, using the IV variance formula, we show that the FEVD SEs for coefficients of both the time-varying and time-invariant variables are uniformly *too small*. In the case of the latter variables, the discrepancy in the FEVD SEs is *unbounded* and grows with the length of the panel and with the variance of the group effects.

Second, using the moment-condition representation, we prove that the coefficients of the time-varying variables in FEVD are exactly the same as in FE. This result is apparent in many of the practical studies that list FE estimates beside FEVD estimates, but it is hardly mentioned in the existing analytical material. An immediate implication is that FEVD estimates, like FE, are *inefficient* if any of the time-varying variables are exogenous.

Third, FEVD usually produces lower variance estimates of time-varying coefficients than HT in small samples. However, it does so by including invalid instruments that produce *inconsistent* estimates. So, even with massive quantities of data those FEVD estimates will deviate from the truth.

Further developments can also be made to the estimator, to exploit the ideas in FEVD while avoiding the problems of that procedure. The advantage of FEVD will be found in smaller samples where the large sample concept of consistency does not dominate. The Monte Carlo simulation studies by Plümper and Troeger (2007a) and Mitze (2009) show a trade-off between bias and efficiency in which FEVD often appears to be better than either FE or HT under quadratic loss. We present Monte Carlo evidence that a standard shrinkage approach combines the desirable small sample properties of FEVD with the desirable large sample properties of the HT estimator, so that it has superior risk to both FEVD and HT over a wide region of the parameter space.

In the next section, we introduce the notation to be used and describe the three-stage FEVD estimator. We summarize the connections between these stages in a theorem, which we prove by comparing the various moment conditions. This approach demonstrates naturally the description of the FEVD estimator as IV. Section 3 compares the correct IV variance formula with the formula implicit in the SEs of the three-stage FEVD approach. The main results are summarized in two further theorems. We also provide an empirical example to illustrate these results and some from the previous section. In Section 4, we examine the relationships between estimators in more detail, allowing the possibility of a trade-off between bias and variance to produce an estimator with lower mean-squared error (MSE). Section 5 reports some Monte Carlo evidence in the spirit of Plümper and Troeger that demonstrates the superiority of a standard shrinkage estimator. Section 6 has some overall conclusions.

2 The Model

The data are ordered so that there are N groups each of T observations. The model for a single scalar observation is

$$y_{it} = X_{it}\beta + Z_i\gamma + u_i + e_{it}, \quad \text{for} \quad i = 1, \ldots, N \quad \text{and} \quad t = 1, \ldots, T. \tag{1}$$

Here, X_{it} is a $k \times 1$ vector of time-varying explanatory variables, and Z_i is a $p \times 1$ vector of time-invariant explanatory variables.[1] The parameters β, γ, the group effect u_i, and the error term e_{it} are all unobserved.

[1]The setup here describes a balanced panel with observations on every t for each i, but the ideas extend to unbalanced panels with more complicated notation. A constant can be represented in this model by including a vector of ones as part of the time-invariant elements, Z.

Some elements of X_{it} or Z_i are correlated with the group effect u_i, in which case we call those variables *endogenous*. Otherwise, we call those variables *exogenous*. With endogenous explanatory variables standard linear regression techniques may produce estimates of the unknown parameters that are inconsistent in the sense that they do not converge to the true parameter values as the sample size grows large. One standard approach to this endogeneity problem is to use the IV technique developed by Hausman and Taylor.

2.1 *Notation*

The presentation is considerably simplified by introducing some projection matrix notation. Let

$$D = I_N \otimes \iota_T, \tag{2}$$

where I_N is an $N \times N$ identity matrix and ι_T is a $T \times 1$ vector of ones. That is, D is a matrix of dummy variables indicating group membership. For any matrix M, we use $P_M = M(M'M)^{-1}M'$ to indicate the projection matrix for M, and we use $Q_M = I - P_M$ to indicate the projection matrix for the null space of M. For example,

$$P_D = D(D'D)^{-1}D' = \frac{1}{T}\left(I_N \otimes \iota_T \iota_T'\right) \tag{3}$$

is the matrix that projects a vector onto D. This particular projection produces a vector of group means. That is, $P_D y = \{\bar{y}_i\} \otimes \iota_T$,where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$. Also,

$$Q_D = I_{NT} - P_D \tag{4}$$

is the matrix that produces the within-group variation. That is, $Q_D y = \{y_{it} - \bar{y}_i\}$ is the $NT \times 1$ vector of within-group differences.

2.2 *The FEVD Estimator*

The FEVD estimator proceeds in three stages, which we detail below. To sharpen the analysis, we assume that the elements of Z are exactly time invariant (not just slowly changing), so that $P_D Z = Z$. An explicit analysis of the slowly changing case yields qualitatively similar insights.

Stage 1: Perform an FE regression of y on the time-varying X. The moment condition corresponding to an FE regression is

$$(y - Xb)' Q_D X = 0. \tag{5}$$

The unexplained component after this first step is $y - Xb$. The group average of the unexplained component is $P_D(y - Xb)$.

Stage 2: Regress the group average of the unexplained component from the first step on the time-invariant Z. The moment condition is $(P_D(y - Xb) - Zg)'Z = 0$. Using the fact that $P_D Z = Z$, this moment condition can be equivalently written as

$$(y - Xb - Zg)'Z = 0. \tag{6}$$

The group-average residuals from this regression are

$$h = P_D(y - Xb - Zg). \tag{7}$$

Stage 3: Regress y on X, Z, and h. The coefficients from this step are the final FEVD estimates. The moment conditions are

$$(y - X\beta - Z\gamma - h\delta)'[X, Z, h] = 0. \tag{8}$$

Theorem 1. The solution for β is b from Stage 1; the solution for γ is g from Stage 2; and the solution for δ is one.

Proof. We need to verify that the moment conditions (8) are satisfied at $\beta = b$, $\gamma = g$, and $\delta = 1$. This requires that

$$(y - Xb - Zg - h)'[X, Z, h] = 0. \tag{9}$$

Substituting in the definition of h from equation (7) and gathering terms, this simplifies to

$$(y - Xb - Zg)' Q_D [X, Z, h] = 0. \tag{10}$$

Using the fact that $Q_D Z = 0$, this further simplifies to

$$(y - Xb)' Q_D [X, Z, h] = 0. \tag{11}$$

The first set of equalities in equation (11) must be satisfied since it is identical to the moment condition (5) that defines b. The second set of equalities must be satisfied since $Q_D Z = 0$. Similarly, the third set of equalities must be satisfied since $Q_D h = 0$, which follows from the definition of h in equation (7) and the fact that $Q_D P_D = 0$. □

2.2 *IV Representation*

Using Theorem 1, we can show that the FEVD estimator can also be expressed as an IV estimator for a particular set of instruments. The major benefit of using the IV representation is that one can draw on a standard toolkit of results. Theorem 1 shows that the FEVD estimates of β are identical to the standard FE estimator b from Stage 1. This estimator is defined by the moment condition (5). Theorem 1 also shows that the FEVD estimates of γ are equivalent to the estimator of g from Stage 2. This estimator is defined by the moment condition (6). Combining both moment conditions, and using the fact that $Q_D Z = 0$, the full moment conditions for the FEVD estimator are

$$(y - X\beta - Z\gamma)'[Q_D X, Z] = 0. \tag{12}$$

In other words, the FEVD estimator is equivalent to an IV estimator using the instruments $Q_D X$ and Z.

3 Variance Formulae

Using standard results for IV estimators, the asymptotically correct sampling variance of the FEVD procedure is

$$V_{\mathrm{IV}}(\beta, \gamma) = (H'W)^{-1} H' \Omega H (W'H)^{-1}, \quad \text{for} \quad H = [Q_D X, Z] \quad \text{and} \quad W = [X, Z]. \tag{13}$$

Here, H is the matrix of instruments and W is the matrix of explanatory variables. Ω is the covariance of the residual, $u_i + \mathrm{e}_{it}$, which can be expressed as

$$\Omega = \sigma_{\mathrm{e}}^2 I_{NT} + \sigma_u^2 I_N \otimes \iota_T \iota_T' = \sigma_{\mathrm{e}}^2 Q_D + \left(\sigma_{\mathrm{e}}^2 + T\sigma_u^2\right) P_D. \tag{14}$$

Using straightforward algebraic manipulation of equation (13), we will later separately expand out the variances of β and of γ for more detailed inspection.

We now compare the correct IV variance formula with the FEVD variance formula. Plümper and Troeger state that the sampling variance of the FEVD estimator can be obtained by applying the standard ordinary least squares (OLS) formula to the Stage 3 regression. Therefore,

$$V_{\mathrm{FEVD}}(\beta, \gamma, \delta) = s^2([X, Z, h]'[X, Z, h])^{-1} = s^2 \begin{pmatrix} X'X & X'Z & X'h \\ Z'X & Z'Z & Z'h \\ h'X & h'Z & h'h \end{pmatrix}^{-1}. \tag{15}$$

Here, $s^2 = \|y-X\beta-Z\gamma-h\|^2 \Big/ dof$, where dof is the degrees of freedom. By application of equation (7), the expression for s^2 can be simplified to

$$s^2 = \|Q_D(y-X\beta)\|^2 \Big/ dof, \tag{16}$$

which we note is the standard textbook FE estimator for σ_{e}^2 when $dof = NT-N-k$ (see, e.g., Wooldridge 2002, p. 271).[2]

Now consider the variance of β. The FEVD variance formula for β is the top-left block of the overall FEVD variance formula in equation (15); using the partitioned-inverse formula this submatrix can be written as

$$V_{\mathrm{FEVD}}(\beta) = s^2\left(X'Q_{[Z,h]}X\right)^{-1}. \tag{17}$$

By expanding out equation (13), the correct variance for β can be written as

$$V_{\mathrm{IV}}(\beta) = \sigma_{\mathrm{e}}^2(X'Q_DX)^{-1}. \tag{18}$$

Note that this is exactly the textbook FE variance formula.

Now we note from equation (16) that s^2 is a consistent estimator of σ_{e}^2. However, the matrices in the FEVD formula (equation 17) and the correct formula (equation 18) differ. The FEVD variance formula for β must therefore be incorrect, and we can show the direction of the error.

Theorem 2. The FEVD variance formula for coefficients on time-varying variables is too small.

Proof. Now $P_D[Z,h] = [Z,h]$, so that $P_D P_{[Z,h]} = P_{[Z,h]}$. Such a relationship between projection matrices implies that $P_D - P_{[Z,h]}$ is positive semi-definite (in matrix shorthand, $P_D \geqslant P_{[Z,h]}$). So, $Q_D \leqslant Q_{[Z,h]}$. That $(X'Q_{[Z,h]}X)^{-1} \leqslant (X'Q_DX)^{-1}$ follows immediately. This inequality will almost always be strict because the $p+1$ variables $[Z,h]$ cannot span the whole of the N-dimensional space of group operator D, and the X's have arbitrary within-group variation. □

The FEVD formula for the variance of β is *biased* in that it systematically *understates* the true sampling variance of the estimator. The essential inequality does not disappear as N gets larger, so the formula is also *inconsistent*. The usual reported SEs will be *too small*.

Now, consider the variance of γ. The FEVD variance formula for γ is the middle block of the overall FEVD variance formula in equation (15). Using an alternative representation of the partitioned inverse, this submatrix can be written as

$$V_{\mathrm{FEVD}}(\gamma) = s^2(Z'Z)^{-1}\left(I+Z'[X,h]\left([X,h]'Q_Z[X,h]\right)^{-1}[X,h]'Z(Z'Z)^{-1}\right). \tag{19}$$

Note that $Z'h = 0$, so that in the partitioned central matrix of the second term only the submatrix corresponding to X will be selected. Then, we have the simplification of equation (19),

$$V_{\mathrm{FEVD}}(\gamma) = s^2(Z'Z)^{-1} + s^2(Z'Z)^{-1}Z'X(X'Q_ZX)^{-1}X'Z(Z'Z)^{-1}. \tag{20}$$

In contrast, by expanding out equation (13), the correct variance for γ can be written as

$$V_{\mathrm{IV}}(\gamma) = \sigma_{\mathrm{e}}^2(Z'Z)^{-1} + T\sigma_u^2(Z'Z)^{-1} + \sigma_{\mathrm{e}}^2(Z'Z)^{-1}Z'X(X'Q_DX)^{-1}X'Z(Z'Z)^{-1}. \tag{21}$$

[2]The usual OLS formula for the SEs from the Stage 3 regression would calculate the scale term using $dof = NT - k - p - 1$, where p is the number of Z variables including the constant and the final minus one allows for the additional regressor h. This divisor would clearly produce an inconsistent estimator of σ_{e}^2 for large N and small T. Plümper and Troeger (2007a, 129) mention briefly an adjustment to the degrees of freedom and, although they do not give an explicit formula, their software employs the divisor $dof = NT - N - k - p + 1$ (Plümper and Troeger 2007b). This adjustment would yield a consistent estimate of σ_{e}^2, but it is nonstandard and slightly biased. To sharpen the subsequent analysis, we use the standard unbiased estimator of σ_{e}^2, in which $dof = NT - N - k$.

Again, s^2 is a consistent estimator of σ_e^2, so the first term in equations (20) and in (21) is essentially the same. However, the expressions are otherwise different, so the FEVD variance formula for γ must also be incorrect. Again, we can show the direction of the error.

Theorem 3. The FEVD variance formula for time-invariant variables is too small.

Proof. As shown in the proof of Theorem 2, $(X'Q_DX)^{-1} \geqslant (X'Q_{[Z,h]}X)^{-1}$ with almost certain strict inequality, so the last term in the FEVD variance formula (equation 20) understates the corresponding term in the correct variance expression (equation 21). The only exception would be the unlikely event that X and Z are exactly orthogonal, causing those terms to vanish. But even then, the FEVD variance formula will be an understatement because it omits the term $T\sigma_u^2(Z'Z)^{-1}$, which must be positive definite whenever there are random group effects. □

In general, the FEVD variance formula for γ is systematically *biased* and *inconsistent*. The usual reported SEs will be *too small*. The extent of the downward bias is unbounded. The correct variance expression includes a term that is directly proportional to the number of observations per group T and to the variance of the group effects σ_u^2. In contrast, the FEVD variance formula, and hence the SEs, are unaffected by these parameters. By increasing either or both these parameters, with everything else held constant, the extent of the downward bias in the FEVD variance formula becomes arbitrarily large.

3.1 *Empirical Example*

Reported results from the applied empirical literature align with these theoretical results. Table 1 presents our replication of Table 1 in Belke and Spies (2008) and shows results for pooled OLS (POLS), FE, FEVD, and HT. We add a column for the results from Stage 2 of FEVD and a row for the coefficient δ that arises in Stage 3 to further illustrate our theoretical results.[3] The first six variables only are shown for brevity. They include logged nominal gross domestic product (GDP) of the importing country *Lngdpim*, logged nominal GDP of the exporting country *Lngdpex*, and logged bilateral real exchange rate *Lrer*, as time-varying variables. The time-invariant variables shown are logged great circle distance in km *Ldist*, border length in km *Border*, and dummy for one or both countries being landlocked *Ll*. Results are estimated from a panel sample of $N = 420$ trading pairs for $T = 14$ years giving 5262 observations.

The coefficients for the first three (time varying) variables are the same for FE and FEVD, as shown by Theorem 1. To illustrate the second aspect of Theorem 1, the coefficients for the next three (time invariant) variables are exactly equal in Stage 2 and FEVD, and the solution for δ is one. Theorem 2 is illustrated by

Table 1 Partial replication of Belke and Spies (2008)

	(1) *POLS*	*(2)* *FE*	*(3)* *Stage 2*	*(4)* *FEVD*	*(5)* *HT*
Lngdpim	0.88	0.68		0.68	0.68
	(0.04)	(0.11)		(0.01)	(0.11)
Lngdpex	0.89	0.71		0.71	0.71
	(0.03)	(0.07)		(0.01)	(0.07)
Lrer	−0.01	0.13		0.13	0.13
	(0.01)	(0.06)		(0.01)	(0.06)
Ldist	−1.27		−1.41	−1.41	−1.75
	(0.11)		(0.04)	(0.00)	(0.16)
Border	−0.00		0.00	0.00	−0.00
	(0.00)		(0.00)	(0.00)	(0.00)
Ll	−0.16		−0.23	−0.23	−0.16
	(0.10)		(0.04)	(0.00)	(0.14)
δ				1.00	
				(0.00)	

Note. Robust SEs are in parentheses.

[3]We are grateful to those authors for supplying their data. We found some occasional small differences in reported SEs, probably due to use of "robust" SEs in the published results. The full table of our replication results is provided in the supplementary materials on the *Political Analysis* Web site.

the way the first three FEVD reported SEs are systematically smaller than the FE ones, in an order of 0.01, 0.01, and 0.01, against 0.11, 0.07, and 0.06, even though the coefficients themselves are identical and the SE formula for FE is well established as being correct under the assumptions of the model.

It is a little harder to illustrate Theorem 3, which says that the FEVD SEs on the time-invariant variables are similarly understated. However, the HT estimator is just indentified in this case, which is the reason the HT coefficients and SEs for time-varying variables are exactly the same as FE. It is no surprise, then, that the coefficient estimates of three time-invariant variables (which are all exogenous) are generally similar for POLS, FEVD, and HT. As expected, the HT SEs are slightly larger but very close to those for POLS, in an order of 0.16, 0.00, and 0.14 against 0.11, 0.00, and 0.10. However, the FEVD SEs are very small, at 0.00 in every case for the precision that is shown. This is most implausible because one would not expect POLS to be generally less efficient, given the structure of this example.

Belke and Spies (2008) is the only paper to our knowledge that reports results for all methods including POLS, FE, FEVD, and HT. Several other applications report both FE and FEVD results (e.g., Krogstrup and Wälti 2008; Caporale et al. 2009; Mitze 2009). In the studies we examined, the FE t statistics were consistently smaller than those reported for FEVD time-varying variables—and often much smaller—except for few cases affected by robust SE formulae. Again, this is despite the fact that the coefficient estimators were actually identical by construction.

4 Comparison to Alternative Estimators

The FEVD estimator was introduced as an alternative to the HT estimator. By also expressing FEVD in its IV representation, we are able to develop insights into their comparative properties. Hausman and Taylor showed that the standard FE estimator is equivalent to an IV estimator with instrument set Q_DX. To that, they add any exogenous elements of X or of Z as further instruments.[4]

To see the relationship more clearly, decompose X and Z into exogenous and potentially endogenous sets: $X = [X_1, X_2]$ and $Z = [Z_1, Z_2]$, where the subscript 1 indicates exogenous variables and the subscript 2 indicates endogenous variables. The HT procedure is then an IV estimator that uses the instrument set $[Q_DX, X_1, Z_1]$. In contrast, the FEVD procedure is an IV estimator that uses the instrument set $[Q_DX, Z_1, Z_2]$.

The first essential difference between these estimators is that the FEVD instrument set *excludes* the exogenous time-varying variables X_1. Of course, X_1 may have no members. In that case, the HT estimator for endogenous Z is not identified, so no useful comparisons can be made.[5] However, if X_1 has known members, then a more efficient estimator than FEVD could be created by augmenting the instrument set with X_1. The second essential difference is that the FEVD instrument set *includes* the potentially endogenous time-invariant variables Z_2. If these variables are in fact correlated with the group effect, then the FEVD estimator is inconsistent.

The FEVD and HT estimators coincide exactly when there are no exogenous elements of X and no endogenous elements of Z.[6] The FEVD procedure is thus primarily of interest when some Z may in fact be endogenous. The essential question raised by Plümper and Troeger is then whether it is better to use a biased and inconsistent but lower variance estimator, or a consistent but higher variance estimator. The question of whether a weak-instruments cure is worse than the disease is a sound one, which has been considered in other contexts by a variety of authors; see for example Bound, Jaeger, and Baker (1995).

Under an MSE loss function, neither the FEVD procedure nor the HT procedure will uniformly dominate the other. MSE can be expressed as variance plus bias squared. Thus, a consistent estimator such as HT will be preferable to the FEVD for sufficiently large sample size.[7] In contrast, for a small sample with a small endogeneity problem, it might be preferable to include the time-invariant endogenous variables Z_2 as instruments, as FEVD does. A more efficient estimator of this type than FEVD would be the IV estimator that augments the set of all valid instruments with Z_2, forming the instrument set $[Q_DX, X_1, Z]$.

[4] Hausman and Taylor describe P_DX as the additional instrument, but this interpretation follows Breusch et al. (1989).

[5] Ideally, one would have theoretical grounds for identifying which elements of X are exogenous. As a practical matter, one could also use an over-identification test to confirm this assumption since the FE estimator of β is consistent.

[6] More precisely, the two estimators are identical when all elements of X are treated as if endogenous and all elements of Z are treated as if exogenous, regardless of the actual endogeneity status.

[7] Of course, consistency does require that valid instruments correlated with Z_2 exist.

One conventional approach to finding a balance would be to select between the competing estimators based on a specification test (Baltagi, Bresson, and Pirotte 2003). If the test rejects the null hypothesis of no difference between estimators, then HT would be selected. Otherwise, the efficient estimator would be selected because the evidence of endogeneity is too weak. Selection of a final estimator based on the results of a preliminary test is known as a *pretest* procedure. Inference based on the SEs of the final selected estimator alone may be misleading; however, bootstrap techniques that include the model selection step can circumvent this problem (Wong 1997).

Since the work of James and Stein (1961), statisticians have understood that shrinking (biasing) an estimator toward a low-variance target can lower the MSE. An extensive literature suggests shrinkage approaches based on using a weighted average of two estimators when one estimator is efficient and the other is consistent; see for example Sawa (1973), Feldstein (1974), Mundlak (1978), Green and Strawderman (1991), Judge and Mittelhammer (2004), or Mittelhammer and Judge (2005). We consider a shrinkage estimator that combines the consistent but inefficient HT estimator and the efficient but possible inconsistent IV estimator. For purposes of illustration, we choose a particularly simple shrinkage approach, but the literature contains many variations on the basic theme, which will have different strengths and weaknesses. If the bias, variance, and covariance of two estimators are known, it is algebraically straightforward to find the weight that minimizes the MSE of a combined estimator. In particular, suppose one estimator ϕ is unbiased. The other estimator χ is biased but has lower variance. The shrinkage estimator then has the form $\chi + w(\phi - \chi)$, where w is the weight placed on the consistent estimator. Straightforward calculus shows that optimal weight that minimizes MSE is

$$w = \frac{\mu_\chi^2 + \sigma_\chi^2 - \sigma_{\chi\phi}}{\mu_\chi^2 + \sigma_\chi^2 + \sigma_\phi^2 - 2\sigma_{\chi\phi}}, \tag{22}$$

where bias is indicated by μ and where variance is indicated by σ.

Of course, the exact bias and variances will usually not be known; however, practical estimates of these terms are readily available for IV estimators. Mittelhammer and Judge (2005) show that plugging in such empirical estimates produces a practical weighted-average estimator. They choose a single w to minimize the sum of MSE over all coefficients. Since we are primarily interested in the MSE of a single coefficient in this analysis, we apply the solution for w, as presented in equation (22) which is the single covariate case of equation 3.5 of Mittelhammer and Judge. We use standard empirical estimates of the variance and covariance terms from application of the basic IV formula (equation 13). The difference between the two estimators provides our estimate of the bias of the efficient estimator since HT is asymptotically unbiased. Mittelhammer and Judge provide detailed discussion on calculating bootstrap percentiles and SEs, through application of Efron's bias–corrected and accelerated bootstrap (Efron 1987). The only change needed for the present context is to account for the panel structure, which is most simply done by resampling at the group level rather than resampling single observations independently.

5 Monte Carlo Evidence

In this section, we compare the practical performance under a range of conditions of various estimators for an endogenous time-invariant Z. In addition to the FEVD and HT estimators, we consider a pretest estimator and a shrinkage estimator. The pretest estimator selects between HT and the IV estimator based on the instrument set $[Q_D X, X_1, Z]$, which treats all Z as exogenous (as FEVD does) in addition to using the HT instruments. The pretest selection is based on the 95% critical value of the Durbin-Wu-Hausman specification test for exogeneity of Z (see, e.g., Davidson and MacKinnon 1993, 237). The shrinkage estimator assigns weights for the two estimators according to a first-stage empirical estimate of the formula in equation 22.

Plümper and Troeger argue for the superiority of the FEVD procedure over the HT approach based on Monte Carlo evidence. Although our simulation design stays close to the original design where appropriate, our design differ from theirs in two fundamental respects.[8] The first difference is that in the Plümper and Troeger Monte Carlo study, the HT estimator was not actually consistent. This is because their data

[8]The authors graciously provided the original simulation code upon request.

generating process had no correlation between X and Z. The fact that the available instruments had, by construction, zero explanatory power for the endogenous variable contrasts sharply with their characterization of the Monte Carlo results (130): "the advantages of the FEVD estimator over the Hausman-Taylor cannot be explained by the poor quality of the instruments." Plümper and Troeger note[11] that the advantage of FEVD persists in their experiments regardless of sample size. However, the asymptotic bias of an IV estimator is the same as the bias of OLS when the instruments are uncorrelated with the endogenous variable and thus irrelevant (Han and Schmidt 2001). In contrast, with a valid and relevant instrument, the bias of the IV estimator will approach zero asymptotically. We therefore consider scenarios in our simulation where the HT estimator is consistent, that is at where at least one instrument for the endogenous Z is valid and relevant.

The second difference is that our simulations account for random variation in the group effect, whereas the Plümper and Troeger code holds the effect (u) fixed across all replications. Mundlak (1978) shows there is no loss of generality in assuming the effect is random because the FE estimator and its related procedures can be described as inference conditional on the realizations of the effect in the sample. Further, the effect needs to be at least potentially random if the relationship between the effect and the regressors is to be described as *correlation*. As Mundlak shows, if the random effect is correlated with the group averages of regressors in unknown ways, then the optimal linear estimator in the random-effects *model* is in fact the FE *estimator*.

The code used by Plümper and Troeger does not simply fix the replicated effects at some sample realization, rather it uses the *Stata* command "corr2data" to fix the sample moments of the variables and the group effects exactly in every replication. The vector of effects is thereby "fixed" by making it exactly orthogonal to the exogenous variables, effectively excluding any practical influence of the group effect in the simulated data. That process does not simulate an FE model but rather one in which there is no group effect at all! By contrast, our random-effects simulation represents the situation where the analyst is uncertain of the magnitudes of the group effects.

We run a series of experiments that vary the degree of endogeneity and strength of instrument. The data generating process for our simulation is

$$y_{it} = 1+0.5x_1+2x_2-1.5x_3-2.5z_1+1.8z_2+3z_3+u_i+\mathrm{e}_{it}. \tag{23}$$

Here, $[x_1, x_2, x_3]$ is a time-varying mean-zero orthonormal design matrix, fixed across all experiments. $[z_1, z_2]$ is a time-invariant mean-zero orthonormal design matrix, fixed across all experiments. z_3 is fixed for all replications in each experiment. z_3 has sample mean zero and variance 1 and is orthogonal to all other variables except x_1. The sample covariance of the group mean of x_1 with z_3 is set exactly to an experiment-specific level, which allows us to vary the strength of the instrument across experiments.[9] The idiosyncratic error term e is standard normal. The random effect u is drawn from a normal distribution in each replication. The expectation of u conditional on z_3 is ρz_3, where ρ works out to be the value of $\mathrm{cov}(z_3, u)$ set in the experimental design. All other variables are uncorrelated with u, and the variance of u conditional on all variables is 1.[10] The level of endogeneity is varied across experiments by changing the value of cov (z_3, u). Each experiment has 1000 replications, which vary the random components u and e. There are 30 groups (N) and 20 periods (T), as reported in Plümper and Troeger (2007a). In implementing the estimators, $[x_1, x_2, z_1, z_2]$ are treated as known exogenous, whereas $[x_3, z_3]$ are treated as potentially endogenous.

Figure 1 illustrates the simulation results for varying instrument strengths and endogeneity levels. The vertical axis in each panel is the square root of MSE of various estimators for the endogenous time-invariant variable z_3. The horizontal axis of each panel is the covariance between the random effect u and z_3. Each panel illustrates different instrument strength, as indicated by stronger instruments having higher correlation between the group-means of x_1 and the endogenous variable z_3. The four panels display the experiments for $\mathrm{corr}(\bar{x}_1, z_3) = 0.15, 0.30, 0.45,$ and 0.60, respectively.[11] Note that, within each panel,

[9]Conditional on a nonzero sample correlation of the endogenous variable and the instrument, the moments of the IV estimator exist, so the Monte Carlo MSE is well defined.
[10]The specified pattern of covariance is implemented through a Choleski decomposition approach.
[11]Because variances of $\bar{x}_1$ and z_3 are both 1, the covariance of these variables equals their correlation.

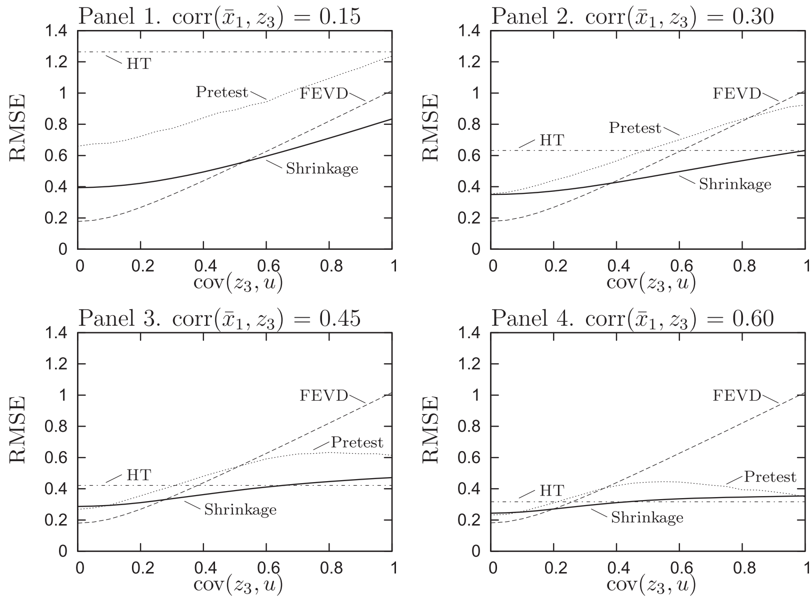

Fig. 1 Performance of the four estimators for varying instrument strengths.

the HT results are unchanging as a consequence of the experimental design. Also, across panels, the FEVD results are unchanging by design.

The most notable feature of Fig. 1 is that neither HT nor FEVD uniformly dominates the other. If reasonably strong instruments are available to implement the HT procedure, and endogeneity is an issue, HT can greatly outperform FEVD as shown in Panel 4 because the higher variance of HT is compensated by lower bias.[12] For all cases when endogeneity is absent (or is mild), FEVD will be the most efficient estimator, as shown at the far left of all panels, because FEVD exploits the true (or approximately true) restriction that z_3 is uncorrelated with u. If the investigator has strong prior reason to believe that endogeneity is not an issue, it makes sense to use that information. Indeed, with informative priors over endogeneity, using a Bayesian procedure that minimizes risk against that prior would be the ideal approach. However, usually, the investigator will be using FE, HT, or FEVD precisely because of concern that endogeneity might be a significant problem.

Rather than relying solely on prior information about the degree of endogeneity, the investigator can rely on evidence from within the data set. Both the shrinkage and the pretest estimators are in this spirit. The shrinkage estimator in particular exhibits remarkably good risk characteristics across all ranges of all four panels, and it clearly dominates the pretest approach under MSE loss. Indeed, the shrinkage estimator often has an MSE lower than *both* the HT and the FEVD estimators, and never is much worse than the better of the two. The Monte Carlo evidence suggests that a shrinkage estimator would almost certainly be the best choice in the absence of prior information that the endogeneity problem is quite small.[13] More generally, if incomplete or uncertain prior information is available, alternatives that explicitly model that information, such as traditional Bayesian techniques or recent variants such as Bayesian model averaging (Hoeting et al. 1999), will likely be the best approach.

[12]The discussion here focuses on the small sample properties. When N is very large, HT will always outperform FEVD if there is endogeneity and valid and relevant instruments exist. For a modest example of relative estimator performance as N grows, see the Appendix, where the case of $N = 300$ and $T = 2$ is illustrated.

[13]Although our focus is on estimator performance, it is worth noting that the Monte Carlo results do confirm that the asymptotic variance formula in equation (13) provides unbiased estimates of the HT and FEVD sampling variance, when σ_e^2 and σ_u^2 are calculated with appropriate degrees of freedom corrections for small sample. Further, the bootstrap quantiles for the shrinkage estimator are reasonably accurate, confirming the results of Mittelhammer and Judge (2005).

6 Conclusions

The FEVD estimator of Plümper and Troeger (2007a) offers the analyst of panel data a way to include time-invariant (and slowly changing) variables in the presence of group effects that are possibly correlated with the explanatory variables. Thus, it appears superior to the existing leading approaches of FE (which omits the time-invariant variables) and Hausman-Taylor (which requires specifying the exogeneity status of each explanatory variable). The motivation of Plümper and Troeger for the procedure was mostly heuristic and their evidence came from Monte Carlo experiments showing that FEVD often displays better MSE properties than both FE and HT. The procedure can be implemented in three easy stages or even more conveniently in the *Stata* package provided by Plümper and Troeger (2007b). This procedure has proved popular with panel data analysts.

Our analytical results and revised Monte Carlo experiments challenge the value of FEVD. Is it still a useful tool?

We find that the coefficients of all the time-varying variables after the three stages of FEVD are exactly the same as FE in the first stage. This fact is sometimes seen in the empirical applications but rarely commented upon with any clarity. Obviously, there is no gain in using FEVD over the simpler FE if these coefficients are the objects of interest. Further, if something is known about the exogeneity of explanatory variables, then these estimates are inefficient because they ignore the extra information. What is worse, unlike the simple first-stage FE, the SEs from FEVD are too small—sometimes very much too small, judging from our empirical example and other published applications. In this case, FEVD is a definite step backward.

The main attraction of FEVD is its ability to estimate coefficients of time-invariant explanatory variables. But, again the third stage is questionable. The same coefficient estimates are given in the second stage, which is a simple regression of the group-averaged residuals from FE on the time-invariant variables. The purported value of the third stage is to correct the SEs, but this reasoning is now known to be false. Indeed, there will be cases where the second-stage SEs—even though they are known to be wrong—will be more accurate than those from the third stage. The example we have provided in Section 3 shows this possibility.

So if FEVD is the label to describe the three-stage procedure, it cannot be recommended for making inferences about any of the coefficients. The coefficient estimator, however, also represents a particular choice of instruments in standard IV. Dropping the three-stage methodology and reverting to an explicit IV approach would allow correct SEs to be obtained in the cases where the estimator is consistent. However, since all the time-invariant variables are used as instruments, the FEVD estimator will be inconsistent if

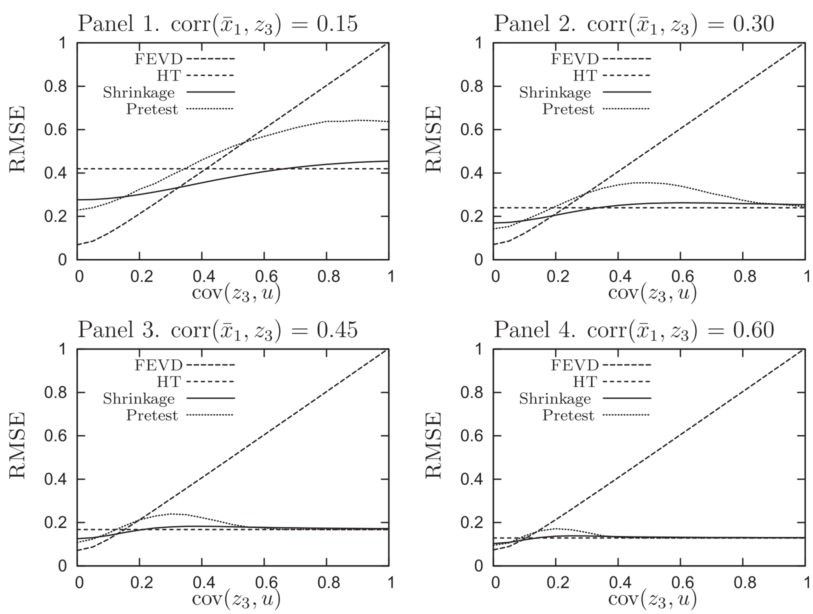

Fig. 2 Relative estimator performance when $N = 300$ and $T = 2$.

any of these are endogenous. The value of this estimator relative to others then depends on the trade-off between inconsistency and inefficiency.

When the objective is reduced MSE, the literature is replete with other methods such as shrinkage estimators known to have good properties. We have provided one such estimator that clearly dominates the FEVD estimator over much of the parameter space and also limits the risk in regions where the FEVD risk is unbounded. We demonstrate the feasibility of such an estimator with SEs found empirically by bootstrapping. In undertaking these investigations, we have also uncovered an explanation for the misleading evidence favoring FEVD that was suggested in the previous Monte Carlo studies.

Appendix

Monte Carlo Results for Large N and Small T

In applications such as labor market studies, the number of groups can be quite large, often in the tens of thousands, since there may be a distinct group for each individual in the study. Figure 2 presents a modest example of the relative behavior of the four estimators as the number of groups grows larger. Each panel in Fig. 2 illustrates the same parameter settings as the corresponding panel in Fig. 1. The simulation code for the figures is identical, except for the N and T settings. Although the overall number of observations is the same in the two figures, the larger number of groups provides more information about the time-invariant variables. Panel 4 illustrates that the relative performance of FEVD can be quite poor for reasonable parameter settings and a modest number of observations.

References

Baltagi, Badi H., Georges Bresson, and Alain Pirotte. 2003. Fixed effects, random effects or Hausman–Taylor? A pretest estimator. *Economics Letters* 79:361–9.

Belke, Ansgar, and Julia Spies. 2008. Enlarging the EMU to the east: What effects on trade? *Empirica* 35:369–89.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90: 443–50.

Breusch, Trevor S., Grayham E. Mizon, and Peter Schmidt. 1989. Efficient estimation using panel data. *Econometrica* 57:695–700.

Caporale, Guglielmo M., Christophe Rault, Robert Sova, and Anamaria Sova. 2009. On the bilateral trade effects of free trade agreements between the EU-15 and the CEEC-4 countries. *Review of World Economics* 145:189–206.

Davidson, Russell, and James G. MacKinnon. 1993. *Estimation and inference in econometrics*. New York: Oxford University Press.

Efron, Bradley. 1987. Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82:171–85.

Feldstein, Martin. 1974. Errors in variables: A consistent estimator with smaller MSE in finite samples. *Journal of the American Statistical Association* 69:990–6.

Green, Edwin J., and William E. Strawderman. 1991. A James-Stein type estimator for combining unbiased and possibly biased estimators. *Journal of the American Statistical Association* 86:1001–6.

Han, Chirok, and Peter Schmidt. 2001. The asymptotic distribution of the instrumental variable estimators when the instruments are not correlated with the regressors. *Economics Letters* 74:61–6.

Hausman, Jerry A., and William E. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica* 49:1377–98.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky. 1999. Bayesian model averaging: A tutorial. *Statistical Science* 14:382–401.

James, W., and Charles Stein. 1961. Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.* Vol. 1, ed. J. Neyman, 361–79. Berkeley, CA; University of California Press.

Judge, George G., and Ron C. Mittelhammer. 2004. A semiparametric basis for combining estimation problems under quadratic loss. *Journal of the American Statistical Association* 99:479–87.

Krogstrup, Signe, and Sébastien Wälti. 2008. Do fiscal rules cause budgetary outcomes? *Public Choice* 136:123–38.

Mittelhammer, Ron C., and George G. Judge. 2005. Combining estimators to improve structural model estimation and inference under quadratic loss. *Journal of Econometrics* 128:1–29.

Mitze, Timo. 2009. *Endogeneity in panel data models with time-varying and time-fixed regressors: to IV or not IV?* Ruhr Economic Paper No. 83.

Mundlak, Yair. 1978. On the pooling of time series and cross section data. *Econometrica* 46:69–85.

Plümper, Thomas, and Vera E. Troeger. 2007a. Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis* 15:124–39.

———. 2007b. xtfevd.ado version 2.00 beta. http://www.polsci.org/pluemper/xtfevd.ado.

Sawa, Takamitsu. 1973. The mean square error of a combined estimator and numerical comparison with the TSLS estimator. *Journal of Econometrics* 1:115–32.

Wong, Ka-fu. 1997. Effects on inference of pretesting the exogeneity of a regressor. *Economics Letters* 56:267–71.

Wooldridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Political Analysis (2011) 19:135–146
doi:10.1093/pan/mpq034

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
e-mail: wgreene@stern.nyu.edu

Plümper and Troeger (2007) propose a three-step procedure for the estimation of a fixed effects (FE) model that, it is claimed, ''provides the most reliable estimates under a wide variety of specifications common to real world data.'' Their fixed effects vector decomposition (FEVD) estimator is startlingly simple, involving three simple steps, each requiring nothing more than ordinary least squares (OLS). 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 FE models, including the estimator of Hausman and Taylor (1981) also (apparently) with impressive gains in efficiency. The article also claims to provide an efficient estimator for parameters on time-invariant variables (TIVs) in the FE model. None of the claims are correct. The FEVD estimator simply reproduces (identically) the linear FE (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 TIVs in an FE model is also incorrect. That part of the parameter vector remains unidentified. The ''estimator'' relies upon a strong assumption that turns the FE model into a type of random effects model.

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 FE 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 ordinary least squares (OLS) without accounting for the common effects, ''works,'' but risks serious omitted variable bias if the FE model (with common effects correlated with the regressors) is appropriate, which is usually the case. A random effects (RE) approach, that is, using generalized least squares (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 FE vector decomposition (FEVD), which 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 FE model, and, along the way solves the problem of non-identification of the coefficients on TIVs 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

Author's note: Helpful discussions of this paper with Neal Beck are gratefully acknowledged. Any remaining errors are, of course, my own.

efficiency gains. An example based on a well-traveled 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 an FE linear regression that contains both time-varying and time-invariant covariates. Using PT's notation,

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

where x_{kit} is a set of K time-varying variables, z_{mi} are M TIVs, 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 $\mathrm{E}[u_i|x_{kit},z_{mi}] \neq 0$. 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

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

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 applies to an unbalanced panel, but at the cost of increased complexity in the notation. In what follows, equation numbers in parentheses, such as (PT-1), refer exactly to the equations in PT, whereas unmodified equation numbers, such as (1), are used for this paper.

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

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, $\mathbf{Z}$, can be absorbed in $\mathbf{X}$. 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 "[t]his 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 TIVs and that is more efficient than the FE model in estimating variables that have very little longitudinal variance. We call this superior alternative 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 $\boldsymbol{\gamma}$ 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 OLS, 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 $\boldsymbol{\alpha}$ by least squares regression of $\mathbf{y}$ on $\mathbf{X}$ and $\mathbf{D}$. As they note (5), ‘‘[w]e 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

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

‘‘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 (PT-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 equations (9–18) in Greene (2008). The transition to their (PT-5),

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

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

Step 2: Based on (PT-5), the estimated unit effects are regressed on $\mathbf{Z}$ 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 (PT-5) and this step. The residuals from the regression are not h_i in (PT-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

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

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

It is suggested that this three-step procedure produces consistent estimators of all the parameters. Step 3 produces the ‘‘correct SEs.’’ ‘‘The third stage also allows researchers to explicitly deal with the dynamics of the TIVs.’’ 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.’’ Our focus at this point is the case of TIVs in $\mathbf{Z}$.

4 Least Squares Algebraic Results

In spite of the extra layer of interpretation in (PT-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 and 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.

The following employs some basic results for partitioned regression in Greene (2008, 27–9). The estimating equation behind the suggested population model in Step 3 is (PT-7). Our empirical counterpart is

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 SE (which is meaningless).

[b] The coefficient estimates on **X** are the original pooled OLS FE 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 SE 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 SEs 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.

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

The $(K + M + 1)$ coefficients $(\mathbf{b}, \mathbf{c}, d)$ are what will be the least squares (FEVD) solutions, not the population parameters. Thus, **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 (PT-4).) The coefficients computed in Step 3 are the OLS solutions in equation (2). Because equation (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 equation (2) to group mean deviations form by premultiplying by $\mathbf{M_D} = \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$.

$$\mathbf{M_D y} = \mathbf{M_D Xb} + \mathbf{M_D Zc} + \mathbf{M_D h}d + \mathbf{M_D w}. \tag{3}$$

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

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

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 equation (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 FE linear regression with a full set of N group dummy variables), the overall constant would be zero.

We now solve for **c** in equation (2). Since **b** is determined, the solution will obey

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

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

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

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

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

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

$$\mathbf{Z}'\mathbf{a}^* = \mathbf{Z}'\mathbf{Z}\mathbf{c}, \tag{8}$$

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

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

$$\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{c}^* = \mathbf{h}d + \mathbf{w}. \tag{9}$$

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

$$\mathbf{h}'(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{c}^*) = \mathbf{h}'\mathbf{h}d + \mathbf{h}'\mathbf{w} \tag{10}$$

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

$$\mathbf{h}'(\mathbf{a}^* + \mathbf{e} - \mathbf{Z}\mathbf{c}^*) = \mathbf{h}'\mathbf{h}d \tag{11}$$

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

Table 2 FEVD three-step estimation (population values in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)
	Step 1—OLS FE		Step 2—OLS, no constant		Step 3—OLS, 1:**X, Z, h**	
Dependent Var.	LWAGESIM		a_i		LWAGESIM	
R^2	.95976		n/a[a]		.95976	
$\mathbf{e}'\mathbf{e}$	79.37106		753.80535		79.37106	
s	0.14927 (0.153221)		1.12841		0.13826 (0.153221)	
Degrees of freedom	3562		592		4152	
	Estimate	SE	Estimate	SE	Estimate	SE
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

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

The derivation does not explain the efficiency payoff. How did the SEs get so small? The appropriate covariance matrix to use for $\mathbf{b}$ is $s^2(\mathbf{X}'\mathbf{M_D}\mathbf{X})^{-1}$. The result at Step 3 is a submatrix of $fs^2(\mathbf{B}'\mathbf{B})^{-1}$, where $\mathbf{B} = (\mathbf{1}, \mathbf{X}, \mathbf{Z}, \mathbf{h})$ and $f = (NT - K - N)/(NT - 1 - K - M - 1)$. 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 $\mathbf{X}'\mathbf{M_D}\mathbf{X}$, 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 $\mathbf{B}'\mathbf{B}$, 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* 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 $fs^2(\mathbf{B}'\mathbf{B})^{-1} << s^2(\mathbf{X}'\mathbf{M_D}\mathbf{X})^{-1}$ 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 $\mathbf{b}_{\text{FEVD}} = \mathbf{b}_{\text{FE}}$.

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 $\boldsymbol{\gamma}$ 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 FE model*. If the focus is shifted to an RE model with TIVs, 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

$\mathbf{x} =$ 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 TIVs are

$\mathbf{z} =$ (fem = gender, blk = race, and ed = education).

The advantage of pivoting the simulation off a real-world 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 $\mathbf{X}$ and $\mathbf{Z}$ from the observed data. But, we will simulate the dependent variable. By this construction, we will make the data conform exactly to the FE 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, $yfitFE_{it,}$ and the estimated residual SD, 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, $\mathbf{c}_{\mathrm{RE}}$. 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, $ysim_{it} = yfitFE_{it} + \mathbf{c}_{\mathrm{RE}}'\mathbf{z}_i + s_{\mathrm{FE}} \times \varepsilon_{it}$, where ε_{it} is a random draw from the standard normal distribution.

Note that the linear regression of $ysim_{it}$ on **X**, **D**, **Z** produces exactly the same coefficients and SEs as the linear regression of $ysim_{it}$ on **X**, **D** because, as noted, **Z** is a linear combination of the columns of **D**, so the least squares estimates of the coefficients on **Z** are zeros.

Thus, $ysim_{it}$ satisfies exactly the assumptions of the model; it is generated by a true FE model with TIVs 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 in column (1). Note that the correct values for the SEs of the FE estimator are also known. Since the disturbances were simulated from a normal distribution with a known SD, s_{FE}, the actual, correct covariance matrix for the FE estimator (conditioned on **X**) is $s_{\mathrm{FE}}{}^2(\mathbf{X}'\mathbf{M_D X})^{-1}$. These true SEs are also shown in parentheses in column (2) of 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 SEs in the third regression compared to those in the first regression, that is, column (6) versus column (2) and in the nonzero coefficients on **Z** in column 5. The SEs have fallen substantially, by factors ranging as high as 6. The population values of $\boldsymbol{\gamma}$ 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 (2007), Mainwaring and Pérez-Liñán (2008), Sova et al. (2009), and Worrall (2008). The striking reappearance of $\mathbf{b}_{\mathrm{FE}}$ in tables of results that present $\mathbf{b}_{\mathrm{FEVD}}$ 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 SEs such as shown in column (6) of Table 2, 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 $\mathbf{b}_{\mathrm{FEVD}}$ and $\mathbf{b}_{\mathrm{LSDV}-}$ 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

$$\mathrm{Est.Var}[\mathbf{b}_{\mathrm{FEVD}}] = s_{\mathrm{FEVD}}{}^2\,(\mathbf{X}'\mathbf{M}_{1,\mathbf{Z},\mathrm{h}}\mathbf{X})^{-1} \tag{12}$$

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

$$\mathrm{Est.Var}[\mathbf{b}_{\mathrm{LSDV}}] = s_{\mathrm{LSDV}}{}^2\,(\mathbf{X}'\mathbf{M}_{\mathrm{D}}\mathbf{X})^{-1}. \tag{13}$$

The scale factor $\mathrm{s}_{\mathrm{FEVD}}{}^2$ is smaller than $\mathrm{s}_{\mathrm{LSDV}}{}^2$ and the former matrix is unambiguously smaller than the latter. The algebraic result is developed after equation (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 is 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.

For the first result, from (PT-1),

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},\ E[\boldsymbol{\varepsilon}|\mathbf{X},\mathbf{Z}] = 0, E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X},\mathbf{Z}] = \sigma^2\mathbf{I}. \tag{14}$$

The LSDV estimator is

$$\mathbf{b}_{\mathrm{LSDV}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_{\mathbf{D}}\boldsymbol{\varepsilon}. \tag{15}$$

It is immaterial whether $\boldsymbol{\alpha}$ is correlated with $\mathbf{Z}$ and $\mathbf{X}$ 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

$$\mathrm{Var}[\mathbf{b}_{\mathrm{LSDV}}|\mathbf{X},\mathbf{Z}] = \sigma^2(\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X})^{-1}. \tag{16}$$

As shown earlier, the FEVD estimator is using $f \times s_{\mathrm{LSDV}}{}^2$ to estimate σ^2 where $f = (NT - \mathrm{N} - K)/(NT - K - M - 2) < 1$. As N increases, f converges to $(T - 1)/T$. That is, the downward bias in the estimator of σ^2 does not go away and is worse the shorter 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 an FE model that contains TIVs, that is, their equation (PT-1). The point that seems to be overlooked in the substantial literature that this proposed estimator has inspired is that in the FE model, if $\boldsymbol{\alpha}$ is assumed to exist *as the set of FE*, then $\boldsymbol{\gamma}$ *does not exist*, so it cannot be estimated, efficiently or otherwise. Reconsider the original model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}. \tag{1}$$

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

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

for some $\boldsymbol{\alpha}^{\mathrm{A}}$. The well-known implication is that it is not possible to estimate $\boldsymbol{\gamma}$ 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 $\mathbf{y}$ on $\mathbf{x}_1$ estimates $(\beta + 2\gamma)$. No other construction is possible. Returning to our equation (1), the solution always employed is to assume $\boldsymbol{\gamma} = \mathbf{0}$ and drop $\mathbf{Z}$ from the model.

The unconvinced reader will now point to PT's

$$\boldsymbol{\alpha} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{h} \tag{PT-6}$$

to argue the opposite. The problem is that (PT-6), like (PT-7), is incorrect. The vector of dummy variable coefficients *in the FE model* is not equal to $\mathbf{Z}\boldsymbol{\gamma}$ plus a disturbance that is uncorrelated with $\mathbf{Z}$. 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 $\mathbf{Z}$. But, this is not an assumption in the FE 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 (PT-6) if it is assumed at the outset that

$$\begin{aligned} y_{it} &= \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + \alpha_i + \epsilon_{it} \\ \alpha_i &= \mathbf{z}_i'\boldsymbol{\gamma} + \eta_i, \end{aligned} \tag{17}$$

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

$$y_{it} = \alpha + \mathrm{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\gamma} + u_i + \varepsilon_{it}, \tag{18}$$

where ε_{it} is as before and now u_i is an RE. 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 $\boldsymbol{\gamma}$. But, that has only been made possible by the additional assumption that the common effects are uncorrelated with the TIVs, an assumption that is not part of the FE specification.

The proposed estimator of $\boldsymbol{\gamma}$ is enabled by assuming that the model is a hybrid of the FE and RE models. The identifying restriction is that η_i in equation (17) 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* 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, 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 OLS. The estimator of the asymptotic covariance matrix based on Step 3 is

$$\mathbf{V}_{(3)} = \frac{\mathbf{e}'\mathbf{e}}{NT-K-M-2}\left(\mathbf{Z}'\mathbf{M}_{\mathbf{1},\mathbf{X},\mathbf{h}}\mathbf{Z}\right)^{-1}, \tag{19}$$

where

$$\mathbf{M}_{\mathbf{1},\mathbf{X},\mathbf{h}} = \mathbf{I} - \mathbf{G}\left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}', \ \mathbf{G} = (\mathbf{1}, \mathbf{X}, \mathbf{h}). \tag{20}$$

For our example, these are the SEs 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

$$\mathbf{V}_{(2)} = \frac{\mathbf{h}'\mathbf{h}}{N-M-1}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}. \tag{21}$$

These would be the SEs 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 SEs 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 equation (17), neither of these matrices is appropriate.

The fixed effect estimator of α_i is given by the result in (PT-4), where $\bar{\mathrm{e}}_i = 0$ for every i and $\boldsymbol{\beta}^{\mathrm{FE}}$ is actually $\mathbf{b}$. Regardless of whether one views the model as the FEM in (1) or the REM in equation (17), 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

$$\begin{aligned} a_i &= \alpha_i + \text{sampling error} \\ &= \alpha_i + v_i, \end{aligned} \tag{22}$$

where the expected value of v_i is 0—since a_i is unbiased. If we now base our interpretation of the model on equation (17), then

$$a_i = \mathbf{z}_i{}'\boldsymbol{\gamma} + \eta_i + v_i \tag{23}$$

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

$$\mathrm{Var}[v_i|\mathbf{X}] = \frac{\sigma_\varepsilon^2}{T} + \bar{\mathbf{x}}_i^{'}\left[\sigma_\varepsilon^2\left(\mathbf{X}^{'}\mathbf{M_D}\mathbf{X}\right)^{-1}\right]\bar{\mathbf{x}}_i. \tag{24}$$

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

$$\mathrm{Var}[a_i|\mathbf{X},\mathbf{Z}] = \sigma_\eta^2 + \sigma_\varepsilon^2\left\{\frac{1}{T} + \bar{\mathbf{x}}_i^{'}\left[\left(\mathbf{X}^{'}\mathbf{M_D}\mathbf{X}\right)^{-1}\right]\bar{\mathbf{x}}_i\right\}. \tag{25}$$

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

$$\mathrm{Asy.Var}[\mathbf{c}^*|\mathbf{X},\mathbf{Z}] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}, \tag{26}$$

based on the second step, where $\boldsymbol{\Omega}$ is a diagonal matrix containing the elements in (25). The matrix computed at Step 3 is irrelevant and bears no resemblance to this or, necessarily, to $\mathbf{V}_{(2)}$. The SEs 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 SEs 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 $\boldsymbol{\gamma}$ 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 LSDV. The claimed efficiency gains under their assumptions are produced by using an erroneous result, equation (PT-7), to motivate an incorrect covariance matrix, both for estimation of $\boldsymbol{\beta}$ and for $\boldsymbol{\gamma}$. The existence of the estimator for $\boldsymbol{\gamma}$ 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 $\boldsymbol{\beta}$ and the asymptotic covariance matrix are correctly estimated at Step 1. The estimator of $\boldsymbol{\gamma}$ coupled with the White (1980) robust covariance matrix obtained at Step 2 is appropriate if the orthogonality assumption is met and is meaningless if not.

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.

References

Akhter, S., and K. Daly. 2009. Finance poverty: Evidence from fixed effect vector decomposition. *Emerging Markets Review* 10:191–206.

Alemán, J. 2008. Labor market deregulation and industrial conflict in new democracies: A cross-national analysis. *Political Studies* 56:830–56.

Brück, T., and H. Peters. 2009. "20 Years of German unification: Evidence on income convergence and heterogeneity," IZA Discussion Paper 4454.

Buckley, J., and M. Schneider. 2007. *Charter schools: Hope or hype?* Princeton, NJ: Princeton University Press.

Caporale, G., C. Rault, R. Sova, and A. Sova. 2009. On the bilateral trade effects of free trade agreements between the EU-15 and the CEEC-4 countries. *Review of World Economics* 145:189–206.

Cornwell, C., and P. Rupert. 1988. Efficient estimation with panel data: An empirical comparison of instrumental variable estimators. *Journal of Applied Econometrics* 3:149–55.

Davis, D. 2009. Price and promotion effects of supermarket mergers. Working Paper. Department of Economics, South Dakota State University.

Greene, W. 2008. *Econometric analysis*. Englewood Cliffs, NJ: Prentice Hall.

Hansen, M. 2007. Using private contracts to create adoptions from foster care. Working Paper No. 2007-03. Department of Economics, American University.

Hausman, J., and W. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica* 49:1377–98.

Mainwaring, S., and A. Pérez-Liñán. 2008. Regime legacies and democratization: Explaining variance in the level of democracy in Latin America, 1978—2004. Political Science Working Paper # 354. Notre Dame, France: Kellogg School.

Plümper, T., and V. Troeger. 2007. Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis* 15:124–39.

Sova, R., L. Albu, I. Stancu, and A. Sova. 2000. Patterns of foreign direct investment in the New EU countries. *Romanian Journal of Economic Forecasting* 2:42–51.

White, H. 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica* 48:817–38.

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

Political Analysis (2011) 19:147–164
doi:10.1093/pan/mpr008

Fixed-Effects Vector Decomposition: Properties, Reliability, and Instruments

Thomas Plümper and Vera E. Troeger
Department of Government, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK
e-mail: tpluem@essex.ac.uk (corresponding author)

This article reinforces our 2007 *Political Analysis* publication in demonstrating that the fixed-effects vector decomposition (FEVD) procedure outperforms any other estimator in estimating models that suffer from the simultaneous presence of time-varying variables correlated with unobserved unit effects and time-invariant variables. We compare the finite-sample properties of FEVD not only to the Hausman-Taylor estimator but also to the pretest estimator and the shrinkage estimator suggested by Breusch, Ward, Nguyen and Kompas (BWNK), and Greene in this symposium. Moreover, we correct the discussion of Greene and BWNK of FEVD's asymptotic and finite-sample properties.

1 Introduction

The fixed-effects (FE) vector decomposition (FEVD) procedure offers a solution to the obvious problem of estimating the effect of time-invariant variables[1] in panel data when at least one time-varying variable is correlated with the unobserved unit effects (henceforth, we refer to variables correlated with the unobserved unit effects as endogenous).[2] In two comments published in this issue, Greene and Breusch, Ward, Nguyen, and Kompas (BWNK) comment on the FEVD procedure. In short, Greene makes the following claims: First, FEVD is an inconsistent estimator. Second, the efficiency gains described in our *Political Analysis* article (Plümper and Troeger 2007) are "illusory." Third, the standard errors (SEs) are too small. In the major part of his article, Greene proofs the obvious no one ever doubted: that for time-varying variables the first and the third stage of FEVD give identical results (which is identical to saying that for time-varying variables FEVD replicates fixed-effects estimates). This is so because time-invariant variables are uncorrelated with the de-meaned variables of the the fixed effects model. Not only is this so evident that Greene's multipage mathematical exercise adds nothing, it is also obviously a substantive property of the fixed effects model and thus beyond criticism. Greene overlooks the fact that the FE model does not generate coefficients for time-invariant variables. Needless to say that FEVD does. Hence, for time-invariant variables, the first stage of FEVD is not identical to the third stage. BWNK make similar claims as the first and third claim of Greene and, in addition and fourth, propose a pretest and a shrinkage estimator, which both try to combine the perceived unbiasedness of Hausman-Taylor (HT) with the efficiency of FEVD.[3]

In this reply, we will show that these claims are either wrong or have become obsolete, as in the case of the SEs issue. Our 2007 *Political Analysis* article already stressed in the title that we are solely interested in the finite-sample properties of FEVD. Despite Greene's persistent inconsistency claims, we remain uninterested in asymptotic properties because infinite properties are not generalizable to finite samples. We will nevertheless show that FEVD is consistent whenever HT is consistent, that is when valid instruments (instruments perfectly uncorrelated with the unobserved unit effects) exist. Greene and BWNK

Authors' note: Supplementary materials for this article are available on the *Political Analysis* Web site.

[1]This symposium deals with the estimation of time-invariant variables in panel data with unit effects. The FEVD procedure also performs better than all known alternatives for estimating variables with low within-variation and time-varying variables uncorrelated with the unobserved unit effects when other time-varying variables of interest are correlated with the unobserved unit effects. For these discussions, see Plümper and Troeger (2007, 2010).

[2]Our definition of endogeneity is thus identical to the one BWNK employ in this symposium.

[3]We use the terminology of King, Keohane, and Verba (1994, 66ff) and Ullah (2004) for bias and efficiency. Note that efficiency solely denotes the sampling variation of an estimator across hypothetical replications.

misrepresent the true properties of FEVD because they ignore FEVD's instrument option. We also demonstrate that—contrary to Greene's repeated claims—there cannot be any doubt that FEVD is more efficient than the FE and HT models (for time-invariant, rarely changing, and exogenous time-varying variables) and less biased than pooled Ordinary Least Squares (OLS) and random effects (for endogenous time-varying variables in finite samples).

In respect to Greene's and BWNK's critique of our variance equation, it is important to note here that our 2007 *Political Analysis* article neither discusses SEs nor does it offer a variance equation. We thus believe that Greene and BWNK criticize the SEs computed by the 2009 version of our *Stata* ado-file, called xtfevd2.0.ado. In a letter to us, Greene claims that he inferred that SEs are wrong because "the standard errors at step three (are) smaller than those at step one" but he denies to have taken notice of the ado-file. Yet, without taking notice of the ado-file, such inferences are odd since our original *Political Analysis* article does not discuss SEs and because OLS estimates come with different types of SE adjustments, for example, Beck and Katz's (1995) panel-corrected SEs. We have replaced the 2.0beta version[4] in early 2010. We will show here that the current xtfevd4.0beta.ado computes SEs based on a variance equation that differs from the different variance equations that Greene and BWNK suggest in their articles and we demonstrate that our variant of FEVD's variance equation computes SEs that are closer to the true sampling variance than the alternative suggestions of both Greene and BWNK.[5]

BWNK accept that a biased but more efficient estimator can be more reliable—that is, can have a smaller root mean squared error—than an unbiased and inefficient estimator and they also agree with us (and thus disagree with Greene) that FEVD has important efficiency gains in comparison to the FE and the HT model. They believe that one can improve on FEVD's small sample properties by merging the procedure with the HT model, which uses time-varying and time-invariant variables assumed to be exogenous as instruments for time-invariant variables assumed to be endogenous. Although we would certainly welcome improvements over FEVD and, more importantly, over the HT estimator, we will show in Section 4 that neither of these models is more reliable than FEVD. Quite to the contrary, FEVD outperforms all currently known models for panel data with endogenous time-varying and time-invariant variables. This is so because the correlation between the time-invariant variables and the unobserved unit effects is unknown and therefore has to be estimated. However, the tests that BWNK's estimators use do not have enough power to render the alternatives to FEVD viable. We will demonstrate that BWNK's claim that their shrinkage model is superior to FEVD results from BWNK's unrealistic assumption of perfectly valid instruments.

This discussion with BWNK has an important element that goes beyond FEVD versus HT versus FEVD/HT variants. Though we accept the data-generating process (DGP) of BWNK, we insist on one exception: BWNK assume that all instruments are perfectly valid, that is in this case, uncorrelated with the unit effects. This assumption does not make sense because applied researchers cannot observe the correlation between potential instruments and the unit effects. This correlation can only be tested, and as we will show, with high imprecision. Restricting the Monte Carlo (MC) analysis to the unobservable case in which instruments are perfectly valid generates entirely unrealistic conditions, which applied researchers do not face. BWNK then show that their shrinkage model performs slightly better than FEVD if and only if instruments are perfectly valid. We show results consistent with theirs that demonstrate that FEVD is vastly superior to the shrinkage estimator in the extremely likely event that the instruments are correlated with the unobserved unit heterogeneity.

2 The Composite Character of the FEVD Procedure

Greene fails to understand that the existence of unobserved unit heterogeneity does not imply a correlation between the true unit heterogeneity and the time-invariant variables. He writes: "But, that has only been made possible by the additional assumption that the common effects are uncorrelated with the TIVs, an

[4]Computer code classified as "beta" version implies that the software is "feature complete," but it does not mean that computer code is perfectly correct. To the contrary, beta versions are meant to be invitations for users to identify bugs and errors and help developing the code further.

[5]Greene (2011) equates efficiency and SEs and talks about "right" and "wrong" variance equations. However, variance equations are attempts to approximate the sampling variation. These approximations are never exactly correct. Therefore, one should talk about better or worse variance equations. In this respect, our current variance equation is better than Greene's but neither is "true."

assumption that is not part of the FE specification.'' He is wrong. The necessity to use a FE model may simply result from a correlation between at least one time-varying variable of interest and the unit effects, whereas all time-invariant variables might be perfectly uncorrelated with the unit effects. Apparently, this setup does not violate the assumptions of the FE specification but is rather a common situation that most applied researchers know only too well and that also motivates the HT model. What makes this estimation situation difficult is the very important fact that the correlation between the time-invariant variable and the unit effects is unobservable (Section 4). Thus, to make FEVD advantageous, three assumptions must be met. First, the DGP ought to be $y_{i,t} = \beta x_{i,t} + \gamma z_i + u_i + \epsilon_{i,t}$ with x being a time-varying and z a time-invariant regressor, u_i denotes the unobserved unit-specific effect with $E\left(u_i | x_{i,t}\right) \neq 0$, that is, the time-varying regressor is endogenous (correlated with the unit effects), and $\epsilon_{i,t}$ is an i.i.d. error term. Although Greene and BWNK assume that z_i and u_i are correlated, we maintain that to applied researchers the correlation between z_i and u_i is unknown. Therefore, the best assumption for the correlation between z_i and u_i is that the correlation is drawn from a probability density function of the correlation between randomly drawn variables.

Ultimately, Greene errs because he maintains a bivariate view on FEVD and merely examines how FEVD differs from FE in estimating time-varying variables. As our 2007 article already clearly states the answer to this question is: nothing. Greene invests astonishing mathematical effort to show that for time-varying variables the first stage of FEVD is identical to its third stage. In proving the obvious and the known, he demonstrates that he does not understand the procedure's simple composite nature. FEVD's purpose is to estimate a coefficient for time-invariant variables and this is where FEVD differs from the FE model that simply does not give an estimate for time-invariant variables because of the perfect collinearity between the estimated FE and the time-invariant variables. In sum, FEVD has characteristics that combine the FE with the pooled OLS model and FEVD analyzes variables that are best analyzed by FE by a *de facto* FE model and variables that are best analyzed by pooled OLS by a *de facto* pooled OLS model. As we concluded in our 2007 *Political Analysis* article, FEVD does better than FE in estimating time-invariant (and rarely changing and exogenous time varying) variables and better than pooled OLS and random effects in estimating endogenous time-varying variables.

Regardless of the true correlation between time-invariant variables (z) and the unobserved unit effects (u), the most difficult problem faced by applied researchers is that this correlation remains unknown and cannot reliably be tested. It is not worth much discussion why the correlation between a variable and the *unobserved* unit effects cannot be *observed*. It clearly requires more to explain why this correlation cannot be estimated reliably and we will do so in Section 4 that responds to BWNK's alternatives to FEVD.

3 Estimator Properties

Debates about the choice of an estimator implicitly or explicitly have to deal with criteria based on which one should evaluate the performance of estimators. Our 2007 *Political Analysis* article already stressed in the title that we are interested in finite-sample properties because asymptotic properties of an estimator do not carry over to finite-sample properties. In this section, we use this discussion to show that FEVD is consistent whenever the HT model is consistent, that is, if and only if the time-invariant variables are uncorrelated with the unit effects or if perfectly valid instruments exist. In a second step, we correct Greene's repeated claim that ''The 'efficiency' gains are illusory.''

3.1 *The Consistency of the FEVD Procedure*

Mainstream econometricians have an awkward way to discuss consistency. They first define a data-generating process, for example, that the time-invariant variables are correlated with the unobserved unit-specific effects. Then they make assumptions, such as the instruments are perfectly valid, that is, uncorrelated with the unit effects. Finally, from this they infer that the instrumental variable (IV) estimator is consistent. This style of discussing consistency is meaningless for applied researchers who want to know which estimator is optimal given the data at hand. It seems a safe bet to argue that applied researchers never have perfectly valid and at the same time strong instruments available. At the very least, these occasions have been very rare. In contrast, we prefer to make the conditions explicit under which estimators are consistent.

When Greene claims that FEVD is inconsistent, he does so with an undertone that suggests it should not be used. Both the explicit and the implicit claims are wrong. FEVD is not necessarily inconsistent and even if it was it could still produce the most accurate point estimates for applied researchers analyzing a limited amount of information. Greene also states that ''It is suggested that this three-step procedure produces consistent estimators of all of the parameters.'' It is not a mere coincidence that he does not provide a quote for this statement. We simply never made this claim.

Rather, we clearly stated in our original article that FEVD is inconsistent if and only if the time-invariant variables are correlated with the unit effects. We write: ''If the time-invariant variables are assumed to be orthogonal to the unobserved unit effects—that is, 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, but this bias is of course just the normal omitted variable bias. Yet, given that the estimated unit effects $\hat{u}$ consist of much more than the unobserved unit effects u and since we cannot disentangle the true elements of u from the between variation of the observed and included variables, researchers necessarily face a choice between using as much information as possible and using an unbiased estimator.'' BWNK almost literally formulate the same statement: ''If these [time-invariant] variables are in fact correlated with the group effect, then the FEVD estimator is inconsistent.'' And even Greene acknowledges in his conclusion: ''The existence of the estimator for $\boldsymbol{\gamma}$ 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.''

However, even this apparent consensus does not reveal the entire truth about FEVD's asymptotic properties. FEVD allows the use of instruments for time-invariant variables in stage 2 and this option renders FEVD consistent whenever HT is consistent. It is actually easy to see why this option makes FEVD consistent when the instruments are perfectly uncorrelated with the unobserved unit effects. The valid instrument simply ensures that the crucial orthogonality assumption between the time-invariant variable and the unit effects that we make in stage 2 is actually satisfied because the valid instrument correctly identifies (parts of) the variation of the time-invariant variable that is uncorrelated with u. As a consequence, the residuals of the second stage include the entire unobserved unit heterogeneity (plus some additional variance if the valid instrument is weak).

Using this option with internal instruments (that is with instruments taken from the set of right-hand-side variables) guarantees that the parameter estimates of FEVD and HT become identical. Accordingly, with internal instruments, both HT and IV-FEVD are consistent if and only if the instrument is perfectly valid. However, in contrast to HT, IV-FEVD allows researchers to use instruments from outside the model. This is an option that HT does not provide for. In the HT model, all possible instruments have to be included as regressors in the estimated model, a ''solution'' that makes the estimation less efficient than IV-FEVD with external instruments.

Table 1 demonstrates, first, that IV-FEVD and HT give identical estimates if an internal instrument is valid and, second, that IV-FEVD is unbiased (and thus consistent) if a valid external instrument exists while the HT estimator is much less reliable in this case because researchers have to include the external instrument into the model causing inefficiency.

Table 1 Bias and efficiency of HT and IV-FEVD with valid internal and external instruments

	Internal instruments		*External instruments*	
Bias and SD beta of z_2	*IV-FEVD*	*HT*	*IV-FEVD*	*HT*
$T = 20, N = 30$	0.886 (1.081)	0.886 (1.081)	0.990 (0.316)	0.858 (0.743)
$T = 20, N = 100$	0.971 (0.241)	0.971 (0.241)	1.001 (0.154)	0.976 (0.233)
$T = 100, N = 30$	0.923 (0.755)	0.923 (0.755)	1.008 (0.294)	0.875 (0.674)

Bias: mean(β) (true beta = 1.0) (values closer to 1.0 indicate less bias). Efficiency: in parentheses: SD of the βs (smaller values indicate higher efficiency). MC setup: DGP: $y = x_1 + x_2 + z_1 + z_2 + u$ + eps, corr(z_2,u) = 0.5, corr(z_2,x_1) = 0.5, corr(z_2,ext. instr.) = 0.5, all other correlations are set to zero, all variables are drawn from a standard normal distribution. See Section 4.4 for a more detailed description of the data-generating process and setup of the Monte Carlo experiments.

Despite all its simplicity, Table 1 reveals two major disadvantages of the HT model. First, although HT (and thus IV-FEVD with internal instruments) is consistent, it is not unbiased in finite samples.[6] Recall that consistency merely requires asymptotic unbiasedness. And second, the HT model requires the inclusion of instruments in the right-hand-side of the model. This unnecessarily leads to inefficiency when valid external instruments exist. In short, the use of external instruments is preferable to the use of internal instruments (unless we have a large N or an extremely large T), and with external instruments, IV-FEVD performs better than HT. Since the set of valid internal and external instruments combined cannot be smaller than the set of valid internal instruments, the set of empirical models for which IV-FEVD is consistent is likely to be larger than the set of empirical models for which HT is consistent. Such a notion, of course, sounds awkward to mainstream econometricians, who are interested in the consistency of estimators rather than in the *ex ante* probability with which estimators give unbiased estimates.[7] In sum, it is not only wrong and misleading to classify FEVD as inconsistent and HT as consistent, FEVD including the instruments option dominates HT.

However, despite FEVD's consistency, we would like to emphasize that the consistency of estimators should not inform applied researchers—at least not unless they have no information about the finite-sample properties of competing estimators. As we have seen in Table 1 with valid internal instruments, consistency is an asymptotic property that does not even guarantee that an estimator is unbiased in finite samples, let alone efficient. One simply cannot generalize from the asymptotic corner solution to the general case of finite information. The asymptotic properties of an estimator do not provide information on its finite-sample properties.[8] Ullah claims: "It is well understood that the use of asymptotic theory results for small and even moderately large samples may give misleading results." (Ullah 2004, ix) The fact that this is well understood does not prevent econometricians and applied researchers from ignoring it.

3.2 *Efficiency*

The imprecision with which our critics use the term consistency is mirrored in Greene's discussion of FEVD's efficiency. In the introduction to his article, he writes: "The efficiency gains are illusory" (Greene 2011, 1), whereas in his conclusion he states: "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)."[9] Since efficiency is defined as a smaller sampling variation, Greene's conclusion contradicts the impression that he seeks to construct in the beginning of his article. When Greene claims that "the 'efficiency' gains are illusory," he is simply wrong. When he writes, FEVD has a smaller sampling variance than HT (and certainly smaller than FE), he is correct. Both claims do not go together.

There is nothing magical and nothing illusory about the fixed-effects vector decomposition procedure: it is more efficient than the FE model for variables defined as "time-invariant" because it uses more information. For these variables, FEVD is marginally less efficient than pooled OLS, whereas for all time-varying variables estimated FEVD's efficiency is identical to that of the FE model. Over all variables, then, FEVD is more efficient than the FE model and less biased than the pooled OLS model, whereas for a single variable FEVD's properties are either identical to FE or pooled OLS.

[6]The bias of HT and IV-FEVD with internal instruments declines rapidly if we increase N but very slowly if T increases. Accordingly, with $N = 100$, we get almost unbiased estimates, with $T = 100$ the bias is not significantly smaller than the bias we see in table 1.

[7]IV-FEVD has an additional advantage over the HT model: it works with a single instrument whereas HT requires two instruments. If those instruments are correlated, the efficiency of HT declines. In those situations, applied researchers would like to exclude one of the instruments—an option which is possible in IV-FEVD but not in the HT model.

[8]Leamer (2010, 31) recently pulled the leg of theoretical econometricians on the consistency issue: "We economists trudge relentlessly toward Asymptopia, where data are unlimited and estimates are consistent, where the laws of large numbers apply perfectly and where the full intricacies of the economy are completely revealed. But it is a frustrating journey since no matter how far we travel, Asymptopia remains infinitely far away."

[9]BWNK almost correctly state: "For all cases where endogeneity is absent (or is mild), FEVD will be the most efficient estimator" (Breusch et al. 2011). We write "almost" because the efficiency is independent of bias. BWNK seem to confuse efficiency and reliability (RMSE) here.

3.3 *Finite-Sample Econometrics*

Finite sample econometrics focuses simultaneously on the bias *and* the efficiency of an estimator. An estimator is biased if the mean of an infinitely large number of repeated estimates of the same model differs from the truth. Since applied researchers do not infinitely resample and then repeat the estimation of their model but rather generate a single-point estimate, bias implies an expected deviation of the estimated coefficient from the truth (the true coefficient). Thus, both bias and inefficiency increase the probability that the point estimate deviates from the truth. The root mean squared error (RMSE) conveniently describes how bias and the sampling variation jointly determine the reliability of a point estimate.

$$\text{RMSE} = \sqrt{\text{var}\left(\hat{\beta}\right) + \text{bias}(\hat{\beta})^2}. \tag{1}$$

An estimator has optimal finite sample properties when it uses all available information and when it is unbiased. Sometimes, this optimal estimator does not exist. In this case, the estimator with the smallest RMSE for the data-generating process mimicking the sample at hand gives estimation results closest to the truth.

4 FEVD, HT, and BWNK's Shrinkage Model

Our 2007 *Political Analysis* article compares the FEVD procedure to fixed effects, random effects, pooled OLS, and the HT model. BWNK accept that the FEVD procedure is more efficient than the HT model and may therefore have superior finite sample properties under identifiable circumstances. However, they also assume that the HT model is less biased whenever the time-invariant variables are correlated with the unit effects. This claim, first, depends crucially on their ignorance of FEVD's instrument option. However, second, it is also wrong. HT is less biased than FEVD if and only if the correlation between the variables used as instruments and the unobserved unit effects is smaller than the correlation between the instrumented time-invariant variables and the unobserved unit effects, which is easy to assume but with currently existing tests impossible to guarantee.

BWNK suggest two estimators that can potentially improve on the HT model and, more importantly, on FEVD. Both estimators seek to combine the relative advantages of both procedures. The first of BWNK's models is a "pretest" estimator. In brief, the procedure computes a variant of the popular Durbin-Wu-Hausman test to decide whether the time-invariant variable is correlated with the unobserved effects and thus whether FEVD or HT are superior and then estimates the appropriate model. The second suggestion is a "shrinkage estimator." This procedure conducts a *de facto* Hausman test to estimate the bias of FEVD and then, depending on the test results, weigh the estimation results of FEVD and HT. Specifically, if the bias of FEVD appears to be large (small), BWNK assign a larger (smaller) weight to HT than to FEVD. The performance of both estimators crucially depends on the power of the test.

Given the relevance of the tests for the performance of the estimators, it seems very puzzling that BWNK's MC analyses do not examine the power of these tests but rather assume that instruments are perfectly valid, that is uncorrelated with the unobserved unit effects. This dubious assumption guarantees the performance advantage of the shrinkage estimator. If we correctly assume that the correlation between time-invariant variables and unobserved unit effects is unknown, then the power of these endogeneity tests determine the relative performance of IV estimators. Due to the poor power of these tests, the performance of IV estimators including of course the shrinkage estimator deteriorates sharply when we abandon the assumption of perfectly valid instruments. With realistic assumptions about the probability density function of correlations between random variables, FEVD outperforms the shrinkage estimator in roughly 95% of the cases in which both estimators compute significantly different results. We will discuss the bias in BWNK's MC design in greater detail before we present some of our MC results and publish the output of additional MC simulations along with the code on our Web site.[10]

[10]http://ww.polsci.org/pluemper/ssc.html.

4.1 *The Pretest Model*

The pretest estimator suggested by BWNK utilizes a Durbin-Wu-Hausman test (DWH test) for the time-invariant variable (Durbin 1954; Wu 1973; Hausman 1978). Assume the data generation process follows

$$y_{it} = \beta_1 x_{it}^1 + \beta_2 x_{it}^2 + \gamma_1 z_i^1 + \gamma_2 z_i^2 + u_i + \varepsilon_{it}, \tag{2}$$

with x being a time-varying variable, z are time-invariant variables, u denotes the unit specific effect, and ε_{it} is the idiosyncratic error term.

BWNK suggest a two-step estimator. In the first step, they estimate an instrumental variable equation, which regresses the endogenous time-invariant variable z^2 on the variables that they assume to be exogenous (x^1, z^1)

$$z_i^2 = \delta_1 x_{it}^1 + \delta_2 z_i^1 + zresid_i. \tag{3}$$

In a second step, they include the residuals ($z$resid) of equation (3) into the original model

$$y_{it} = \beta_1 x_{it}^1 + \beta_2 x_{it}^2 + \gamma_1 z_i^1 + \gamma_2 z_i^2 + \phi z\text{resid}_i + \xi_{it}. \tag{4}$$

BWNK (and the DWH test) assume that if ϕ is significant z^2 is indeed endogenous to the time-invariant part of the error term u.

However, this test is reliable if and only if the instruments are perfectly valid. The test rapidly loses power if the instruments are slightly correlated with the unobserved unit effects. Yet, whether the instruments are correlated with the unobserved unit effects is as unknown to the applied researcher as whether the time-invariant variable of interest covaries with the unit effects. Therefore, the probability of selecting an instrument that has a higher correlation with the unit effects than the instrumented variable is as high as the desired opposite choice. Thus, even in the unlikely case that valid instruments exist, researchers would find it hard to tell which variable is in fact exogenous. In other words, the test presupposes information that applied researchers cannot have. As a consequence, the "test" does not solve the all-important problem of deciding whether the instruments are (close enough to) perfectly valid. Since it fails to solve this problem, it also cannot answer the question whether and to which extent time-invariant variables are correlated with the unit effects. Testing for an unknown correlation by assuming another unknown correlation to be known, is logically inconsistent.[11]

4.2 *The Shrinkage Model*

In addition to the pretest estimator, BWNK suggest a shrinkage estimator that aims at combining FEVD's efficiency and HT's assumed unbiasedness. The shrinkage estimator uses a weighted average of FEVD and HT so that shrinkage = FEVD + w(HT − FEVD). Shrinkage estimators have repeatedly been used in situations where one pure estimator is less biased and another more efficient. Therefore, BWNK suggest a standard solution that has worked elsewhere. But can it work when the correlation with the unobserved unit effects is unknown?

The quality of shrinkage estimators for a specific estimation problem depends on whether a larger weight is placed on the estimator that gives more reliable results. BWNK hold that if bias, variance, and covariance of two estimators are known, it is straightforward to find a weight that minimizes the MSE of the combined estimator. They use the following weight

$$w = \frac{\mu_{\text{FEVD}}^2 + \sigma_{\text{FEVD}}^2 - \sigma_{\text{FEVD,HT}}}{\mu_{\text{FEVD}}^2 + \sigma_{\text{FEVD}}^2 + \sigma_{\text{HT}}^2 - 2\sigma_{\text{FEVD,HT}}}, \tag{5}$$

[11] Appendix A in the supplementary data demonstrates the absence of power of this "test." It shows that the test is valid if and only if the instruments are both valid and strong. If we make correct assumptions about the probability density distribution of correlations between random variables the test gives far more "false positives" than "correct positives."

where μ stands for bias, σ^2 for variance, and $\sigma_{\text{FEVD,HT}}$ for the covariance between HT and FEVD. BWNK use empirical estimates for the variance and covariance that are readily available from the IV variance equation.

In order to make the shrinkage estimator work, BWNK make what they admit to be a problematic assumption: the point estimates of HT are more accurate than the point estimates of FEVD. Based on this assumption the bias of FEVD is computed as the difference between the HT and FEVD point estimates. This assumption is indeed problematic: Without perfectly valid instruments, both HT and FEVD are biased. If the correlation between the instruments and the unobserved unit effects exceeds the correlation between the time-invariant variable and the unit effects, HT is more biased than FEVD. In addition, since FEVD is under all circumstances more efficient than HT, and since inefficiency leads to an expected larger deviation from the truth, HT's point estimate will in many cases be further away from the truth than FEVD's point estimate. For these reasons, the test on which BWNK base the computation of the weight, and especially the measure for bias μ cannot work properly.

Most importantly, the Hausman test does indicate "endogeneity" not only when time-invariant variables are endogenous but also when the instrument is poorly chosen. If the correlation between the instruments and the unobserved unit effects exceeds the correlation between the time-invariant variable and the unobserved unit effects, the Hausman test detects a significant difference in the estimation result and the weighting formula will put a larger weight on the HT model, despite the fact that it is less efficient and more biased.

4.3 *The Dubious Power of Unrealistic Assumptions*

We have seen that the performance of the pretest and the shrinkage estimator depends on the validity of instruments, which—as we have repeatedly said—remains unknown to the applied researcher. Before one chooses such an estimator, one would certainly want to know how reliable the estimation results are when researchers make wrong assumptions about the validity of instruments or when perfectly valid instruments do not exist.

However, this is exactly what BWNK assume away. BWNK's MC design not only makes the extremely strong assumptions that all instruments are perfectly valid, they also hide this important assumption behind jargon that one can only understand in case one perfectly understands the HT model. It makes sense to cite BWNK (2011) to understand the setup of the simulations they use: "Here, $[x_1, x_2, x_3]$ is a time-varying mean-zero orthonormal design matrix, fixed across all experiments. $[z_1, z_2]$ is a time-invariant mean-zero orthonormal design matrix, fixed across all experiments. z_3 is fixed for all replications in each experiment. z_3 has sample mean zero and variance 1 and is orthogonal to all other variables except x_1. The sample covariance of the group mean of x_1 with z_3 is set exactly to an experiment-specific level, which allows us to vary the strength of the instrument across experiments. The idiosyncratic error term e is standard normal. The random effect u is drawn from a normal distribution in each replication. The expectation of u conditional on z_3 is ρz_3, where ρ works out to be the value of $\mathrm{cov}(z_3,u)$ set in the experimental design. All other variables are uncorrelated with u, and the variance of u conditional on all variables is 1."

It is advisable to read this passage twice to seek to understand how BWNK introduce the all-important assumption that instruments are perfectly valid. In fact, this assumption can be found in the fairly innocent sounding statement that "all other variables are uncorrelated with u." This assumption only sounds innocent. Recall that the HT model utilizes the exogenous time-varying and time-invariant variables as instruments for the endogenous variables. Thus, BWNK assume that all instruments are perfectly uncorrelated with the unobserved unit effects throughout their simulations. Given that this correlation remains unknown in reality, this assumption resembles divine revelation.[12]

4.4 *MC Experiments without Arbitrarily Truncated Correlation Space*

MC simulations should mimic the conditions faced by applied researchers. Since applied researchers cannot know the true correlations between the unobserved unit effects and both the time-invariant variables

[12]Observe that the "perfectly valid instruments" assumption also influences the power of the pseudo-Hausman test between FEVD and HT estimates.

and their potential instruments, MC analyses should not be restricted to the unlikely case where instruments are perfectly valid. Rather, potential tests of these correlations have to become substantial parts of comparing the relative performance of estimators. Three correlations (and the number of observations) influence the relative performance of FEVD, HT, the pretest model, and the shrinkage estimator:

1. The correlation between the time-invariant variable and the unit effects, corr(z,u). This correlation cannot be observed.
2. The correlation between the instruments and the unit effects, corr(m,u). This correlation cannot be observed.
3. The correlation between the instruments and the time-invariant variable, corr(m,z). This correlation can be observed.

We report the results of two different sets of MC experiments here. In the first set, we fix the correlations between the time-invariant variables and the unobserved unit effects (endogeneity), between the instruments and the unobserved unit effects (validity), and between the instruments and the time-invariant variables (strengths) at various levels in each case. In the second set of experiments, we randomly draw the correlations from an approximation of the probability density functions of correlations between random variables.

Since we find that the pretest estimator performs poorly, we do not report results here (but we include them in the replication material). Likewise, the HT model is dominated by a combination of FEVD and the shrinkage estimator. In those few cases in which FEVD performs worse than HT, HT does worse than the shrinkage estimator. Therefore, we also do not report HT results and focus on a comparison between FEVD and the shrinkage estimator that—according to BWNK—performs under all conditions better than both FEVD and HT. We will demonstrate that this claim results solely from the arbitrary restrictions on the correlation space they impose on their MC analyses, which are therefore highly misleading and should be interpreted with caution as we have explained before.

For the MC experiments, we follow the setup in our 2007 *Political Analysis* article and the MC setup used by BWNK and assume the following data-generating process:

$$y_{it} = \beta_1 x_{it}^1 + \beta_2 x_{it}^2 + \gamma_1 z_i^1 + \gamma_2 z_i^2 + u_i + \varepsilon_{it}, \tag{6}$$

where x^1 and x^2 are time-varying variables and z^1 and z^2 are time-invariant explanatory variables. Note that we include two time-varying and two time-invariant variables because this is required by the HT model. FEVD just needs a single time varying plus a single time-invariant variable. Thus, in order to make a comparison between FEVD on the one hand, and HT, the pretest, and the shrinkage estimator on the other hand viable, we need to include two time-varying and two time-invariant variables into the model. x^2 and z^1 follow an orthonormal design matrix and are irrelevant for discussion of the estimation of time-invariant variables. Note that deviations from this assumption increase the RMSE of HT, the pretest, and the shrinkage estimator, but not of FEVD. We also accept BWNK's MC design in drawing all variables as well as the idiosyncratic error term ϵ and the unit-specific effect u from a standard normal distribution.

We are solely interested in the reliability of the coefficient for z^2, which is the time-invariant variable potentially correlated with the unit-specific effects u. The unit mean of x^1 serves as instrument for z^2 (we will call the instrument m below and repeatedly claim that instruments are time invariant, which of course holds for instruments which are the unit means of time-varying variables). Again following BWNK's design, we vary the strength of the instrument by changing the correlation between the unit mean of x^1 and z^2. In addition, we also vary the correlation between the unit mean of $x1$ and u in order to manipulate the validity of the instruments. All coefficients are set to 1.[13]

The only difference then between BWNK's MC design and ours is that we replace BWNK's unrealistic assumption that the instrument is perfectly uncorrelated with the unobserved unit effects by the realistic

[13]The number of units N equals 30 and the number of periods T equals 20; results for different combinations of N and T are available from our Web site.

assumption that this correlation can vary between 0 and 1, is unknown, and thus has to be estimated using the tests that BWNK use for their pretest and shrinkage estimators.

4.4.1 MC experiment 1: fixed correlations

Since negative correlations are functionally equivalent to positive correlations, we restrict the possibility space for each correlation to values between 0 and 1. It is important to note that the three correlations that matter here are not independent of each other. If corr(z,u) = 1 and corr(m,z) = 1, then corr(m,u) must be 1 as well. This limits the possibility space in a relevant way because the optimal IV estimation requires that the instrument m is highly correlated with the time-invariant variable z and uncorrelated with the unit effects u. Yet, there are limits to the strengths of an instrument. This matters because the advantage of an IV estimator becomes maximal if the corr(z,u) is very high and corr(m,u) zero. However, such a constellation is only possible within limits. For example, when corr(z,u) = 0.8 and corr(m,z) = 0.7, then corr(m,u) = 0 is impossible as the correlation matrix between the three variables becomes singular. Ironically, when instruments are most valuable, they cannot be simultaneously perfectly valid and strong.

Table 2 displays differences in reliability (RMSE) of FEVD versus BWNK's shrinkage estimator for different levels of instrument strengths corr(m,z) and severity of the endogeneity problem corr(z,u). Choosing five different levels of correlations suffices since all competing estimators are well behaved and are not prone to erratic changes of the RMSEs. For example, when the correlation between the time-invariant variable z_2 and the unit effects increases, the RMSE cannot become smaller but will increases unless instruments used are perfectly valid in which case it will stay roughly constant. Likewise, if the correlation between instruments used and the unobserved unit heterogeneity becomes larger, the RMSE increases.

We subtract the root mean error of the shrinkage estimator from the RMSE of FEVD, so that negative values imply superiority of FEVD, positive values imply superiority of the shrinkage estimator. Missings result from combinations of correlations that lead to a singular correlation matrix.

The gray-shaded cells indicate constellations in which FEVD gives more reliable estimates than BWNK's shrinkage estimator. The table demonstrates that when we relax BWNK's assumption of perfectly valid instruments, FEVD outperforms the shrinkage estimator. Our simulations clearly indicate that BWNK's claim that the shrinkage estimator combines the advantages of FEVD and HT is wrong. This is so because their shrinkage estimator makes the wrong assumption that whenever the point estimate of FEVD and HT differ, HT must be closer to the truth. This assumption is based on the asymptotic properties that do not carry over to finite samples and on the assumption that researchers know the correlation between instruments and the unobserved unit effects.

4.4.2 MC experiment 2: random draw of correlations from an approximation of the probability density function of correlations of random variables

The above tables may wrongly give the impression that FEVD is about twice as reliable on average as the shrinkage estimator. This impression is misleading because it depends on the implicit assumption that all correlations occur equally likely. However, correlations between random variables are not uniformly distributed. Rather, correlations close to zero are much more likely to occur than large correlations. In fact, the distribution of the correlation coefficient resembles a truncated normal distribution when the number of observations exceeds 20. Since negative and positive correlations are functionally identical for our discussion of instruments, the probability density function for random variables resembles a half normal distribution truncated at 1.0 with a standard deviation (SD) of around 0.25 when $N = 20$ and smaller as N increases. We assume a realistic SD of 0.25.

Figures 1a and 1b depict the relative advantage of FEVD versus the shrinkage estimator for repeated draws of corr(z_2,u) and corr(m,u) from the probability density distribution of correlations between random variables.[14] Note that figure 1a displays only models in which FEVD's estimates are significantly closer to the truth than the estimates of the shrinkage estimator while figure 1b displays models in which the shrinkage estimator has a significant advantage over FEVD. To make a "significant" improvement, an estimate must be 0.5 closer to the truth than the alternative estimate (the true coefficient is fixed at 1.0).

[14]Since the strength of the instrument corr(m,z) can be observed, we set this value to 0.5.

Table 2 Difference in the RMSE between FEVD and shrinkage estimator

		Strengths of instrument: corr(m,z₂)				
		0	0.1	*0.3*	*0.5*	*0.7*
When corr(m,u) = 0 (perfect validity)						
Endogeneity of z_2: corr(z_2,u)	0.00	−0.159	−0.152	−0.137	−0.112	−0.041
	0.10	−0.152	−0.159	−0.130	−0.082	−0.030
	0.30	−0.089	−0.073	−0.034	0.040	0.081
	0.50	−0.037	−0.007	0.113	0.207	0.289
	0.70	−0.002	0.054	0.283	0.410	0.488
	0.90	−0.002	0.072	0.459	—	—
When corr(m,u) = 0.1						
Endogeneity of z_2: corr(z_2,u)	0.00	−0.210	−0.188	−0.164	−0.125	−0.065
	0.10	−0.178	−0.186	−0.172	−0.100	−0.050
	0.30	−0.134	−0.144	−0.087	−0.009	0.058
	0.50	−0.057	−0.045	0.013	0.144	0.228
	0.70	−0.028	−0.034	0.143	0.338	0.444
	0.90	0.005	−0.012	0.303	—	—
When corr(m,u) = 0.3						
Endogeneity of z_2: corr(z_2,u)	0.00	−0.475	−0.446	−0.439	−0.346	−0.240
	0.10	−0.435	−0.466	−0.403	−0.322	−0.213
	0.30	−0.361	−0.393	−0.359	−0.231	−0.094
	0.50	−0.237	−0.288	−0.244	−0.074	0.049
	0.70	−0.135	−0.233	−0.159	0.075	0.238
	0.90	−0.038	−0.182	−0.054	—	—
When corr(m,u) = 0.5						
Endogeneity of z_2: corr(z_2,u)	0.00	−0.869	−0.901	−0.856	−0.692	−0.491
	0.10	−0.845	−0.885	−0.877	−0.712	−0.460
	0.30	−0.703	−0.804	−0.760	−0.569	−0.350
	0.50	−0.596	−0.637	−0.672	−0.401	−0.182
	0.70	−0.412	−0.531	−0.528	−0.240	−0.010
	0.90	—	−0.487	−0.432	—	—
When corr(m,u) = 0.7						
Endogeneity of z_2: corr(z_2,u)	0.00	−1.299	−1.331	−1.295	−1.086	−0.776
	0.10	−1.258	−1.295	−1.308	−1.060	-0.761
	0.30	−1.165	−1.216	−1.227	−0.927	−0.633
	0.50	−0.971	−1.084	−1.080	−0.767	−0.452
	0.70	−0.852	−0.925	−0.954	−0.619	−0.278
	0.90	—	—	—	—	—

Missings imply combinations of correlations that lead to a singular correlation matrix.

These two figures speak for themselves. In figure 1a, we plot all combinations of corr(u,z) and corr(m,z) for which FEVD estimates are significantly more reliable than shrinkage estimator estimates. Figure 1b displays all combinations in which the shrinkage estimator performs significantly better than FEVD. As expected, we see that shrinkage does better if and only if the correlation between z and u is high (z is highly endogenous) and the instruments are close to perfectly valid. What matters more, however, is that of the 5000 models that we estimated with $N = 30$ and $T = 20$, FEVD was significantly better in 741 cases while the shrinkage estimator was significantly better in only 37 cases (for the remaining cases, there is no significant difference between the two estimators). Likewise, when $N = 100$ and $T = 20$, FEVD performed significantly better in 773 cases, whereas the shrinkage model outperformed FEVD in 21 cases. And when we used $T = 100$ and $N = 20$, the result was 710:38 in favor of FEVD. In other words, if we just look at constellations in which FEVD and the shrinkage estimator produce significantly different estimates, FEVD is more reliable in 95.2 ($N = 30$, $T = 20$), 97.3 ($N = 100$, $T = 20$), and 94.9 ($N = 20$, $T = 100$) percent of the cases. Of course, when we increase T to infinity, the shrinkage estimator will eventually become increasingly competitive and with a huge number of periods eventually

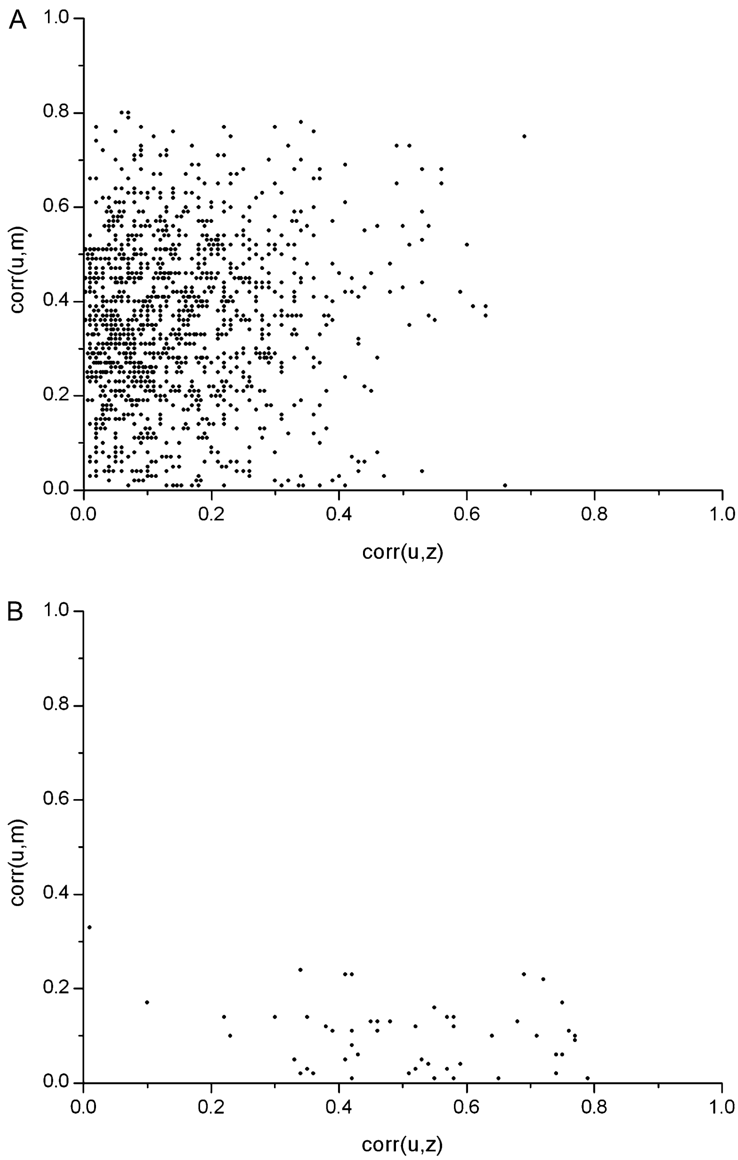

Fig. 1 (a) Constellations in which FEVD significantly outperforms BWNK's shrinkage model ($N = 30$, $T = 20$). (b) Constellations in which BWNK's shrinkage model significantly outperforms FEVD ($N = 30$, $T = 20$).

outperforms FEVD. However, the advantage of FEVD declines extremely slowly. With $N = 20$ and $T = 500$, FEVD performs better in 94.4% of the cases in which the estimation results of FEVD and the shrinkage estimator differed significantly. Although econometricians prefer to assume that asymptotic properties carry over to finite samples if only N or T is large enough (mainstream econometricians seem to think that large enough means roughly the data set at hand), our results suggest that Asymptotia begins nowhere near the size of commonly used data sets.[15] Presumably, not a single researcher with the usual budget constraint would ever collect high frequency data over the time necessary to ensure that the shrinkage estimator does better than FEVD—especially when estimators with sufficiently nice finite sample properties (such as FEVD) are available.

[15]Of course, we would really like to know at which T the shrinkage estimator becomes more reliable than FEVD. However, we had a strict deadline for writing this reply and since simulations with large T are time consuming, we could not identify the threshold at which the relative performance advantage of FEVD disappears. We will report additional MC output on our Web page http://www.polsci.org/pluemper.

4.5 *IV-FEVD*

FEVD not only yields more reliable estimates than HT and the shrinkage estimator in the absence of a reliable endogeneity test for the correlation between time-invariant variables and the unobserved unit effects. The instrument option that both Greene and BWNK ignore improves the performance of FEVD whenever the shrinkage estimator does better than FEVD. In fact, whenever the shrinkage estimator performs better than FEVD, the IV option ensures that FEVD, call it IV-FEVD, does better than the shrinkage estimator. Of course, this situation requires that the correlation between time-invariant variables (including instruments such as the between variation of time-varying variables) and the unobserved unit effects becomes known due to the invention of a reliable test. So let us assume for the sake of argument that in the near or far future some bright scholar develops a test that reliably detects the correlation between time-invariant variables and the unobserved unit effects. Should researchers use the HT or possibly BWNK's shrinkage model?

The simple answer is: neither. Two strategies are superior: First, if someone invents a reliable test for estimating the correlation between time-invariant variables and unobserved unit effects (if such a test is possible at all), applied researchers should use the instrumental variable variant of FEVD because under plausible conditions IV-FEVD is more efficient than HT and allows the use of external instruments. And second, if such a test continues to be inexistent, FEVD is on average far more reliable than any currently existing alternative including most notably the HT model and BWNK's preferred shrinkage estimator. In sum, FEVD dominates the HT and the shrinkage model because the instrumental option of FEVD guarantees that our procedure works at least as good as HT and is likely to be better. This is so because researchers can optimize the instrumental equation without having to change and potentially to spoil the model. In fact, among all possible ways to use instruments, the HT model offers the least elegant and least efficient way. Thus, FEVD is not only more reliable than BWNK's shrinkage model, it also allows applied researchers to use instruments in stage 2, an option that both BWNK and Greene entirely ignore. The use of valid instruments does not only make FEVD consistent, the use of instruments also guarantees the superiority of IV-FEVD over HT in situations when the correlation between z and u is high and instruments are perfectly valid or at least close to perfectly valid.

4.6 *Summary and Discussion*

BWNK's claim about the superiority of the shrinkage estimator is—to borrow Greene's favorite adjective—illusory. Indeed, the shrinkage estimator's proclaimed superiority depends on the illusion that applied researchers are able to identify perfectly valid instruments. Once we relax this assumption, the shrinkage estimator's advantage vaporizes. In fact, we have shown that if we make roughly correct assumptions about the probability density function of the correlation of random variables, FEVD outperforms the shrinkage estimator as it gives more reliable results in approximately 95% of the cases in which the estimation results of FEVD and the shrinkage estimator differ significantly.

We are especially puzzled by the fact that theoretical econometricians apparently like to make the assumption that the correlation between instruments and the unobserved unit effects are likely to be very low when they at the same time assume that the correlation between time-invariant variables and the unobserved unit effect is high. These inconsistent assumptions evidently imply that econometricians like to amplify a potential problem (endogeneity) while they minimize the problems associated with instruments. However, there is no logical difference between instruments and instrumented variables. All that distinguishes these variables is an arbitrary decision of a researcher to call some time-invariant variables m for instruments while the other time-invariant variables are called z.

5 Standard Errors

Standard errors are commonly interpreted as the uncertainty of a point estimate. Standard errors should be identical to the sampling distribution, which is the distribution of point estimates researchers would get if nature would repeatedly draw errors from a standard normal distribution. While it is of course possible to simulate these repeated draws in MC analyses, applied researchers have to live with what nature gives them: a single draw. As a consequence, the sampling distribution cannot be observed, but only approximated.[16]

Interestingly, BWNK, Greene, and the fevd4.0beta suggest different approximations for the computation of SEs.

The xtfevd.beta4.0.ado variance formula is

$$V_{\mathrm{FEVD4.0}}(\beta,\gamma)=(H'W)^{-1}H'\Omega H(W'H)^{-1} \tag{7}$$

$$H=[\ddot{X},Z] \tag{8}$$

$$W=[X,Z] \tag{9}$$

$$\Omega=\sigma_{\epsilon}^{2}I_{NT}+\sigma_{\hat{\eta}}^{2}I_{N}\otimes\iota_{T}\iota_{T}^{'}, \tag{10}$$

where $\ddot{X}=x_{it}-\frac{1}{T}\sum_{t=1}^{T}x_{it}$ (x demeaned) and where I_N is an $N\times N$ identity matrix, ι_T is a $T\times 1$ vector of ones, $\sigma_{\hat{\eta}}^{2}$ stands for the variance of the residuals (eta) of the second stage regression of the FEVD procedure, the unexplained part of the unit specific effects, whereas $\sigma_{\hat{u}}^{2}$ indicates the variance of the estimated unit specific effects of the first stage fixed effects regression.

Breusch et al. suggest an approximation which is similar at the first glance, but which gives very different SEs. Replacing equation (10) by

$$\Omega_{\mathrm{BWNK}}=\sigma_{\epsilon}^{2}I_{NT}+\sigma_{\hat{u}}^{2}I_{N}\otimes\iota_{T}\iota_{T}^{'}, \tag{11}$$

they get much larger SEs.

In a different approach, Greene proposes

$$V_{\mathrm{GREEN}E}(\gamma)=(Z'Z)^{-1}Z'\Omega Z(Z'Z)^{-1} \tag{12}$$

$$\Omega_{\mathrm{GREENE}}=\sigma_{\eta}^{2}+\sigma_{\epsilon}^{2}\left\{\frac{1}{T}+\bar{x}_{i}^{'}[X'\ddot{X}]^{-1}\bar{x}_{i}\right\}, \tag{13}$$

We use the standard way to compare the performance of these three different attempts to approximate the true sampling variation: MC analyses (see, e.g., Beck and Katz 1995). Following their example, we define underconfidence as

$$\text{underconfidence}=100\frac{\sqrt{\sum_{k=1}^{K}\left(\mathrm{SE}(\hat{\beta}^{k})\right)^{2}}}{\sqrt{\sum_{k=1}^{K}\left(\hat{\beta}^{k}-\bar{\hat{\beta}}\right)^{2}}} \tag{14}$$

Since we agree with Greene and BWNK that the SEs of the time-varying variables are just the fixed-effects SEs, we only show the SEs of the time-invariant variables z.

Observe, first, that OLS is overconfident (this was the main reason for why xtfevd2.0beta was overconfident), with computed SEs being much smaller than the sampling distribution. On the other end of the spectrum, BWNK's computed SEs are too large, leading to underconfidence. Both Greene's and xtfevd4.0beta SEs are fairly accurate, with Greene's performing better when both N and T are small (too small to pool) and ours being more accurate when N and T are above 20. More generally, in both cases, the accuracy of SEs depends largely on T, suggesting that estimates of pooled data with a T smaller than 20 or 25 are problematic.

[16]BWNK and Greene claim that their (different!) approximations are "correct" (BWNK) and "appropriate" (Greene).

Table 3 Under/overconfidence as a function of N and T

SD(u) = 1; corr(x,z,u) = 0 Number of observations		Mean (SE(z))/SD(beta(z))			
N	*T*	*FEVD4.0*	*BWNK*	*GREENE*	*Pooled OLS*
	10	91	121	102	56
10	30	86	124	92	37
	50	83	124	87	29
	70	82	121	85	25
	100	83	126	85	22
	10	108	137	119	59
30	30	98	133	103	37
	50	100	139	103	31
	70	103	146	106	28
	100	96	138	98	22
	10	106	135	117	58
50	30	99	134	104	37
	50	99	137	102	30
	70	100	140	103	26
	100	99	140	101	22
	10	111	139	122	59
70	30	107	145	112	40
	50	105	145	109	31
	70	100	139	102	26
	100	98	138	100	22
	10	112	140	123	60
100	30	104	140	109	38
	50	99	137	103	30
	70	104	144	107	27
	100	100	140	102	22

Although Table 3 assumes that no regressor is correlated with the unit effects, Table 4 repeats the MC exercise for two levels of correlation between u and x and z on the one hand and for different standard deviations of u.

Of course, these results repeat what we have already reported before: pooled OLS is overconfident and BWNK's variance equation underconfident. However, when the variance of u relative to the variance of all other variables goes to infinity—that is when we analyze an almost pure cross-section in which the R^2 approaches zero—BWNK's variance formula improves, but of course, this does not look like a correctly specified model. Note that in these cases both Greene's and our suggestions for the computation of SEs become marginally overconfident.

Finally, Table 5 varies the correlation between u and both x and z. In general, the higher the correlation between the regressors (x, z) and the unobserved unit effects (u) the larger the accuracy gap between the fevd4.0beta SEs and the Greene formula. Though the differences remain small, we conclude that for all models in which $T > 20$, the fevd4.0beta SEs are closer to the true sampling distribution than Greene's SEs. At the same time, both fevd4.0beta and Greene's variance formulas are superior to the pooled OLS, BWNK's, and also fevd2.0beta formula. Thus, we may concede that Greene improved over the FEVD variance equation that existed when he wrote his article. Yet, the variance formula implemented in fevd4.0beta is more accurate than Greene's suggestion.[17]

[17]The accuracy gap between our variance equation and Greene's increase further when we estimate variables with low within and large between variation as "invariant."

Table 4 Under/overconfidence as a function of var(u) with constant var(x): var(z)

$SD(u)$		*Mean(SE(z))/SD(beta(z))*			
		FEVD	*BWNK*	*GREENE*	*Pooled OLS*
0.5	N = 20	104	183	118	58
1	T = 30	98	137	104	38
1.5	corr(x,u) = 0	92	113	94	28
2	corr(z,u) = 0	86	101	88	23
3		92	102	93	21
4		88	96	89	19
5		89	95	89	18
0.5	N = 20	103	204	115	55
1	T = 30	94	160	99	35
1.5	corr(x,u) = 0.5	93	143	95	27
2	corr(z,u) = 0.5	92	134	93	23
3		87	119	88	19
4		94	124	94	19
5		93	120	93	19

6 Conclusion

In this article, we respond to our critics and reinforce the case for using FEVD when researchers are simultaneously interested in time-varying variables correlated with the unit effects and time-invariant variables. Briefly, our main arguments can be summarized as follows. First, what appears to be Greene's over-riding criterion for estimator evaluation—consistency—should not inform the choice of estimators for the typical sample sizes analyzed by applied researchers in general and especially not for the estimation problems discussed in this symposium. Infinite sample properties of estimators cannot and thus should not be generalized to finite samples. Still, FEVD has reasonable properties not just for finite samples but also in Asymptotia: Because of the instrument option, FEVD is consistent when valid internal or external instruments exist.

Second, Greene and BWNK make the correct point that SEs were too small in a previous beta version of our xtfevd-ado file. However, we have corrected this defect in the currently available version and long before their manuscripts have been accepted by *Political Analysis*. While Greene's suggested alternative variance equation produces SEs which are closer to the sampling variation than the ones from our previous ado-file (even this does not hold for BWNK's variance equation), the currently available version of our ado-file (xtfevd4.0beta.ado) generates SEs that are even closer to the true sampling variation. The gap between the accuracy of our and Greene's variance equation widens if we consider rarely changing variables. BWNK's variance equation generates underconfident SEs.

Third, Greene corrects his claim that FEVD's efficiency gains are illusory in the conclusion of his article. We do not have anything to add here.

Fourth, Greene's proof that the first stage of FEVD is identical to its third stage when variables are time varying has always been obvious and never was in doubt. FEVD differs from the FE model in respect to time-invariant variables and only in respect to them, see Plümper and Troeger (2007) for further discussion.

Fifth, the shrinkage estimator proposed by BWNK outperforms FEVD (without the IV option) if and only if instruments are simultaneously very strongly correlated with the assumed endogenous variables and almost uncorrelated with the unobserved unit effects. However, not only are these conditions extremely unlikely to exist, but the assumed weak correlation or absence of correlation between the instruments and the unit effects is exactly that: assumed, that is, it cannot be either observed or reliably tested. We show that with realistic assumptions about the correlations between random variables, FEVD is far more reliable than the shrinkage estimator. In addition, even under conditions in which the proposed shrinkage estimator would outperform FEVD without the IV option, FEVD with the IV option will always

Table 5 Under/overconfidence as function of corr(x,u) and corr(z,u)

		Mean(SE(z))/SD(beta(z))			
corr(x,u)	*corr(z,u)*	*FEVD*	*BWNK*	*GREENE*	*Pooled OLS*
N; T = 20; 30; SD(u) = 1					
0	0.1	96	138	103	41
	0.3	94	155	105	42
	0.5	97	176	107	45
	0.7	103	206	107	48
	0.9	112	318	131	69
0.1	0.1	94	130	99	37
	0.3	99	147	101	38
	0.5	98	167	101	40
	0.7	96	215	110	48
	0.9	112	320	127	66
0.3	0.1	97	125	96	33
	0.3	98	136	95	33
	0.5	96	161	99	37
	0.7	104	202	103	42
	0.9	110	317	122	61
0.5	0.1	95	121	96	31
	0.3	92	136	97	32
	0.5	102	164	102	36
	0.7	100	208	107	41
	0.9	101	304	114	55
0.7	0.1	93	115	92	28
	0.3	93	128	92	29
	0.5	97	153	96	32
	0.7	98	195	101	36
	0.9	106	302	111	51
0.9	0.1	98	115	92	28
	0.3	98	128	92	29
	0.5	102	153	96	32
	0.7	99	195	101	36
	0.9	108	302	111	51

outperform the shrinkage estimator. Researchers are thus never better off using the shrinkage estimator than using FEVD either with or without the instrument option. The shrinkage estimator does not improve upon FEVD under realistic assumptions.

We think the relevance of our arguments go beyond the narrow debate about FEVD. The mainstream econometricians' practice to generalize from asymptotic properties to finite sample properties remains a genuine problem, which in many instances leads to unnecessarily poor estimation results in applied research. Inferences based on an analysis with a consistent estimator with poor finite sample properties are worse than inferences based on an analysis using an inconsistent estimator with better finite sample properties if the sample size is finite. This is a known and recurrent theme for readers of *Political Analysis*.[18]

This article also sheds some light on the extent to which mainstream econometricians use unrealistic assumptions to support the estimation procedure they suggest. BWNK's assumption of perfectly valid instruments is not uncommon in econometrics. However, this assumption lies directly at odds with

[18]See most recently Gawande and Li (2009): "The infinite-sample properties (e.g., consistency) used to justify the use of estimators like 2SLS are on thin ground because these estimators have poor small-sample properties. (. . .) They may suffer from excessive bias and/or Type I error."

the notably poor performance of endogeneity tests for the correlation between time-invariant variables and unobserved unit effects. Tests should not have to presuppose what they pretend to test.

References

Amemiya, Takeshi, and Thomas E. MaCurdy. 1986. Instrumental-variable estimation of an error-components model. *Econometrica* 54:869–81.

Beck, Nathaniel, and Jonathan Katz. 1995. What to do (and not to do) with time-series cross-section data. *American Political Science Review* 89:634–47.

Breusch, Trevor, Michael B. Ward, Hoa Thi Minh Nguyen, and Tom Kompas. 2011. On the fixed-effects vector decomposition. *Political Analysis* 19:165–9.

Durbin, James. 1954. Errors in variables. *Review of the International Statistical Institute* 22:23–32.

Gawande, Kishore, and Hui Li. 2009. Dealing with weak instruments: An application to the protection for sale model. *Political Analysis* 17:236–60.

Greene, William. 2011. Fixed-Effects vector decomposition: A magical solution to the problem of time-invariant variables in fixed effects models? *Political Analysis* 19:135–46.

Hausman, Jerry A. 1978. Specification tests in econometrics. *Econometrica* 46:1251–71.

Hausman, Jerry A., and William E. Taylor. 1981. Panel data and unobservable individual effects. *Econometrica* 49:1377–98.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. Designing Social Inquiry: Scientific Inference in Qualitative Research. Princeton: Princeton University Press.

Leamer, Edward E. 2010. Tantalus on the road to Asymptotia. *Journal of Economic Perspectives* 24:31–46.

Mundlak, Yair. 1978. On the pooling of time series and cross section data. *Econometrica* 46:69–85.

Plümper, Thomas, and Vera E. Troeger. 2007. Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit effects. *Political Analysis* 15:124–39.

———. 2010. *Efficient estimation of exogenous variables in panel data with endogenous variables*. Unpublished manuscript, University of Essex.

Ullah, Aman. 2004. *Finite sample econometrics*. Oxford: Oxford University Press.

Wooldridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Wu, De-Min. 1973. Alternative tests of independence between stochastic regressors and disturbances. *Econometrica* 41:733–50.

Political Analysis (2011) 19:165–169
doi:10.1093/pan/mpr012

FEVD: Just IV or Just Mistaken?

Trevor Breusch, Michael B. Ward, Hoa Thi Minh Nguyen, and Tom Kompas
Crawford School of Economics and Government, The Australian National University, Canberra, ACT 0200, Australia
e-mail: trevor.breusch@anu.edu.au (corresponding author), michael.ward@anu.edu.au, hoa.nguyen@anu.edu.au, tom.kompas@anu.edu.au

Fixed effects vector decomposition (FEVD) is simply an instrumental variables (IV) estimator with a particular choice of instruments and a special case of the well-known Hausman-Taylor IV procedure. Plümper and Troeger (PT) now acknowledge this point and disown the three-stage procedure that previously defined FEVD. Their old recipe for SEs, which has regrettably been used in dozens of published research papers, produces dramatic overconfidence in the estimates. Again PT concede the point and now adopt the standard IV formula for SEs. Knowing that FEVD is an application of IV also has the benefit of focusing attention on the choice of instruments. Now it seems PT claim that the FEVD instruments are always the best choice, on the grounds that one cannot know whether any potential instrument is correlated with the unit effect. One could just as readily make the same specious claim about other estimators, such as ordinary least squares, and support it with similar Monte Carlo assumptions and evidence.

1 Introduction

Thomas Plümper and Vera Troeger in this issue attempt to defend the fixed effects vector decomposition (FEVD) estimator as introduced in *Political Analysis* in 2007. This statistical procedure for models with both time series and cross-section dimensions has proved popular with researchers in many fields where such data structures arise. The motivation for FEVD was mostly heuristic, with evidence from Monte Carlo experiments in which FEVD appeared to display better mean-squared error properties than other estimators. The two critiques of FEVD in this issue, one by William Greene (2011) and the other by us, instead provide formal analyses of the FEVD estimator and its SE recipe. Although the approach we used in our critique differs from Greene's, our findings are consistent with his on all points where the analyses overlap.

The response by Plümper and Troeger (2011) does not find any error in these critiques. Instead of offering formal analysis, they again rely on simulation experiments based on *ad hoc* assumptions, without precise definitions of the conditions of the debate and without mathematics to lend precision to the discussion. In this rejoinder, we do not attempt to address every issue in the wide-ranging response. However, we have a duty to inform those who might otherwise rely on empirical results obtained with FEVD and those who might be attracted to use the method in the future.

2 Model

The model in vector form is:

$$y = X\beta + Z\gamma + u + e. \tag{1}$$

The Xs are time-varying explanatory variables; the Zs are time-invariant explanatory variables; there is an unobserved group or unit effect u which is also time invariant and an overall unobserved error e that varies in both dimensions. It is possible that u might be correlated with some X or Z variables, in which case the variables so affected are described as endogenous, otherwise they are exogenous.

3 Instrumental Variables

The most transparent definition of the FEVD coefficient estimator is as linear instrumental variables (IV) with instruments $[Q_D X, Z]$. Here Q_D is the projection matrix that converts a data vector into deviations from group means. Other familiar estimators also have descriptions as IV estimators: fixed effects (FE) uses instruments $Q_D X$, and pooled OLS uses instruments $[X, Z]$. The Hausman-Taylor (HT) estimator is another, although it requires the variables to be partitioned, so the instruments are $[Q_D X, X_1, Z_1]$, where subscript "1" refers to assumed exogenous and "2" refers to assumed endogenous. Relative to HT, the FEVD instrument set omits the exogenous time-varying variables X_1 but includes the endogenous time-invariant variables Z_2.

FEVD and HT coincide exactly if both $X_1$and Z_2 are empty. Put differently, FEVD is the HT estimator under the assumption that all the time-varying Xs are possibly endogenous and all the time-invariant Zs are exogenous.[1] There is nothing exceptional or objectionable about the coefficient estimates produced by FEVD: they are perfectly sensible IV estimates under a particular exogeneity assumption. The difference between HT and FEVD, as estimation *strategies*, is that HT would employ the FEVD instruments under an explicit exogeneity assumption, whereas FEVD would use a predetermined set of instruments as a "canned" solution, without regard to any such reasoning. On this view, PT's contribution is to recommend a standard estimator in what is likely an inappropriate context.[2]

The IV interpretation of FEVD also simplifies life for the applied researcher. The internal *Stata* commands *ivregress* and *xtivreg* can calculate the same coefficient estimates as FEVD—and they provide appropriate SEs.[3]

4 Variances and SEs

Plümper and Troeger (2011) state that "our original PA article does not discuss SEs." That is patently untrue. In Plümper and Troeger (2007), FEVD is defined by a three-stage regression procedure, where the sole purpose of the third stage is to adjust the SEs. The coefficients produced by the first two stages remain unchanged, so the final stage is irrelevant in estimating the coefficients. The need for the third stage to obtain correct SEs is discussed in multiple places in Plümper and Troeger (2007), starting with the abstract.[4] Clearly the recipe in Plümper and Troeger (2007) is to use the SEs produced by the last stage, perhaps after adjustments for residual correlations and degrees of freedom.

We showed that this published recipe provides SEs that are dramatically too small, leading to overconfidence in the results. We wish to be perfectly clear on one point: our analysis was of the algorithm in the published paper. Through an abundance of caution, we also confirmed that the same defective recipe is implemented in the version of their software *xtfevd.ado* that was in distribution up until early 2010. The method of calculating SEs is the issue here, not a particular software implementation.[5]

Plümper and Troeger (2011) now disown their three-stage procedure as a source of SEs. Instead, they adopt a standard IV approach, exactly as recommended in our equations (13) and (14). However, now they claim a substantive disagreement confined to the appropriate estimate of the variance of the

[1]The additional step in the HT estimator of weighting by an estimate of the covariance matrix of the errors is redundant here, because with this instrument set the estimates are just identified.

[2]One aspect of Plümper and Troeger (2007) that is largely overlooked in the subsequent debate is the treatment of "rarely changing" explanatory variables, that is, variables that have little within-group variability over time, in other words Xs that are close to being Zs. As far as we can tell, the proposal is to treat such Xs as if they were in fact Zs. Thus, even though as Xs they are thought to be endogenous, they will nevertheless be used directly as instruments, rather than using only their within-group variation $Q_D X$. That proposal is an obvious extension of the result in the literature on "weak instruments" in simple models that OLS might in some cases be a better estimator than IV, even when the former is inconsistent. The survey by Murray (2006) is a readable introduction to this, now extensive, literature.

[3]The internal *Stata* command for the HT estimator called *xthtaylor* reports an error when the field for listing the X_1 variables is left empty (at least up to *Stata* Release 11). There is no logical reason for that restriction, in that the estimator is still well defined provided the Z_2 field is similarly empty, although there may be practical reasons in the implementation of the computer code.

[4]On page 125, they say more explicitly: "This third stage allows computing correct SEs for the coefficients of the (almost) invariant variables." Further discussion of SEs occurs on page 129: "The estimation of stage 3 proves necessary for various reasons. First of all, only the third stage allows obtaining the correct SEs. Not correcting the degrees of freedom leads to a potentially serious underestimation of SEs and overconfidence in the results." An even more detailed and explicit discussion of these third-stage SEs is presented in the working paper version publicly posted by the authors (Plümper and Troeger 2006).

[5]PT incorrectly attribute the error as a software issue, and then claim our analysis is "obsolete" since they have updated the software. They further disclaim responsibility for the software, stating "beta versions are meant to be invitations for users to identify bugs and errors and help developing the code further" (Plümper and Troeger 2011, footnote 4). While the software faithfully implemented the written recipe, this declaration of *Caveat Emptor* will nonetheless surprise many of those applied workers who used this code to produce published research. The only indication of the status of the software as "beta" is that single word located in the header comment that is embedded in the code itself—a level of detail that most users would overlook and a use of programming jargon that few would understand. We note that the current code (version 4.0, Plümper and Troeger, 2010) is also described as beta in the same way. Moreover, neither this code itself nor the associated help file contains any notice of the previously erroneous SEs—almost a year after PT first removed version 2 from circulation, while citing our critique as the cause.

unit effect, σ_u^2. Moreover, they devote much of Section 5 to discussing their Monte Carlo evidence that our recommendation is underconfident. Since PT provide no definition of the crucial symbols $\hat{\eta}$ and $\hat{u}$ in their equations (10) and (11), respectively, we can only turn to their software implementation to deduce the meaning. There we discover that $\sigma_{\hat{u}}^2$, in the formula attributed to us, is an estimate of the variance of $Z\gamma+u$ in equation (1) above. This is clearly inconsistent with our definition of σ_u^2, which is the variance of u alone. Moreover, we find that $\sigma_{\hat{\eta}}^2$ in their code is an estimate of the variance of u alone. In short, while PT rightly criticize a nonsensical variance estimator, they wrongly attribute it to us. At the same time, they adopt exactly the approach we recommend, while calling it their own.

5 Exogeneity Assumptions

The many claims of superiority of FEVD over other estimators that are made in Plümper and Troeger (2007), and which remarkably are expanded upon in Plümper and Troeger (2011), become clearer through the analytical lens we provide. The FEVD coefficient estimator is just linear IV with a particular predetermined instrument set. Pooled OLS and FE are also IV but with different instruments. There are infinitely many other IV estimators, including HT that requires other input of exogeneity assumptions to specify subsets of the explanatory variables.

Of course, the simplicity of FEVD is attractive to the applied worker: a canned or packaged solution needs no effort to justify its exogeneity assumptions. But the same simplicity is provided by OLS, so simplicity alone is not reassuring. The properties of any particular IV estimator depend on the validity and relevance of its instruments as measured, respectively, by being uncorrelated with the error term and strongly correlated with the endogenous explanatory variables. For any particular choice of instruments, it is always possible to invent Monte Carlo simulation experiments in which your chosen IV estimator does well, or where it does badly compared to a different choice of instruments. Other procedures that choose instruments or mix estimators based on empirical evidence will be less efficient when the predetermined instrument choice is nearly optimal, but might be safer bets when the prejudice is badly wrong. None of these statements relies on the asymptotic theory being a fully reliable guide to actual performance in small samples.

Remarkably, Plümper and Troeger (2011) argue at length that it is impossible to make informed instrument choices in a world where everything depends on everything else in ways we do not fully know. Such postmodernism will surely surprise the mainstream reader of *Political Analysis*. It is hard to see how it fits with the business of Plümper and Troeger (2007), which purports to be parameter estimation in causal statistical models. Parameters only have meaning if, at least in principle, we can identify them independently from the melange of potential influences in our model of the world. If you believe that everything-depends-on-everything-else in ways you cannot fathom, then it is impossible to talk meaningfully about parameters—let alone to estimate them. On the other hand, with appropriate theory and other knowledge of the phenomenon being studied, we can describe causes to which we can give names, and which with suitable data we can perhaps quantify.

In specifying statistical models, we routinely use theory, other knowledge, and experience to determine the variables of interest and to interpret the parameters of the relationship. Judgments of exogeneity and specifications of suitable instruments come from the same sources. We use certain instruments because we can reason about their likely efficacy, based on knowledge of the particular phenomenon being studied. Of course, this knowledge will always be imperfect, and mistakes will be made, but if you are not willing to make some claims about exogeneity you are not entitled to make claims about causal inference either.

The arguments used by PT can easily be driven to absurd extremes. They claim for instance that since you cannot observe u, you cannot know if an instrument candidate is exogenous. We could with equal justification argue that since you cannot observe e, you cannot know if Q_DX is exogenous. (After all, you can just as easily have a time-varying omitted variable as a time-invariant one.) In that case, if the solution to lack of knowledge of exogeneity is to take the simplest predetermined set of instruments, regardless of the implications, simple pooled OLS will always be the preferred estimator!

6 The "Instrument Option" of IV-FEVD

Despite arguing that HT and other targeted IV estimators are illogical, on the grounds that exogeneity status is unknowable, Plümper and Troeger (2011) highlight an additional "instrument option," whereby targeted IV estimation is subsumed into the definition of FEVD:

> FEVD allows the use of instruments for time-invariant variables in stage 2 and this option renders FEVD consistent whenever HT is consistent. . . . Using this option with internal instruments (that is with instruments taken from the set of right-hand-side variables) guarantees that the parameter estimates of FEVD and HT become identical. . . . However, in contrast to HT, IV-FEVD allows researchers to use instruments from outside the model. This is an option that HT does not provide for. (Plümper and Troeger 2011)

Thus, it is claimed, not only is FEVD as good as HT when the information needed for HT is available (indeed, it is claimed the estimates are identical), but when further information is available FEVD can make gains that HT cannot achieve because HT is limited to instruments based on the set of explanatory variables. In passing, we note that this supposed limit on HT is strange and contrived: the HT estimator is explicitly introduced as an IV technique in a paper that also discusses outside instruments.

The claims that IV-FEVD will replicate HT or better its efficiency are dubious on other grounds. The method described here applies IV to the second-stage regression (of the group means of residuals from FE onto the time-invariant *Z*s). The coefficients for the time-varying *X*s will remain simply FE from the first stage, so they will not benefit from the additional instruments and therefore cannot be more efficient than HT. The coefficients of the *Z*s from IV estimation of the second stage will in general be different from IV estimation of the full model containing both *X*s and *Z*s because there will no longer be the neat orthogonality among the explanatory variables and instruments that allows (original) FEVD to be interpreted as joint IV on the full model. Moreover, if the instruments over-identify the model, there will be further efficiencies to be exploited in the full model by weighting by the covariance matrix of the errors, as is done in the HT estimator. Without an explicit proof being given, the presumption must be that IV on the partial model of stage two will make less than fully efficient use of the instruments.

7 Conclusions

Plümper and Troeger respond to their critics and attempt to reinvent FEVD so that it transcends the criticism. They do not identify any errors in the analyses of the critics, but instead continue to assert the superiority of their invention. They acknowledge at the end of their response that the SEs calculated using the recipe of Plümper and Troeger (2007) are too small, and they now adopt the IV approach to obtaining correct SEs. In the process, the three-stage regression procedure—which ironically *defines* FEVD for many users—has been abandoned. But their claim that the FEVD instruments are always optimal is specious because the same claim can just as logically be made for simple OLS and just as easily supported with *ad hoc* Monte Carlo assumptions.

The situation where the IV estimator that corresponds to FEVD will be of interest is where the time-varying *X*s are all possibly endogenous, and will be treated as such, but no similar fears are held regarding the time-invariant *Z*s. Indeed, at the end of their response PT seem to concede as much:

> In this paper, we respond to our critics and reinforce the case for using FEVD when researchers are simultaneously interested in time-varying variables correlated with the unit effects and time-invariant variables. (Plümper and Troeger 2011)

Of course, this is precisely the exogeneity assumption under which the more complicated HT method reduces to FEVD. The estimator in this case is simple linear IV, so standard ideas will motivate the estimator and indicate its statistical properties, and standard software is available to compute the estimates and provide appropriate standard errors. Extensions such as the ‘‘rarely changing variables’’ problem of Plümper and Troeger (2007) are also familiar in the IV literature. Thus, there is no need for ‘‘vector decomposition’’ and no distinct estimator to call FEVD. With appropriate justification for the instruments, FEVD is just IV. Without that justification, FEVD is just mistaken.

References

Greene, William. 2011. Fixed effects vector decomposition: a magical solution to the problem of time invariant variables in fixed effects models? *Political Analysis* 19:135–46.

Murray, Michael P. 2006. Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives* 20:111–32.

Plümper, Thomas, and Vera Troeger. 2006. *Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects.* Discussion paper, Department of Government, University of Essex, Version tirc_80, August 24, 2006.

———. 2007. Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis* 15:124–39.
———. 2010. xtfevd.ado version 4.0 beta. http://www.polsci.org/pluemper/xtfevd.ado.
———. 2011. Fixed effects vector decomposition: properties, reliability and instruments. *Political Analysis* 19:147–64.

Political Analysis (2011) 19:170–172
doi:10.1093/pan/mpr011

Reply to Rejoinder by Plümper and Troeger

William Greene
Department of Economics, Stern School of Business, New York University, New York, NY 10012
e-mail: wgreene@stern.nyu.edu

The remarks in my paper, ''Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time Invariant Variables in Fixed Effects Models?,'' were a comment on the econometric methods proposed in Plümper and Troeger, ''Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects,'' *Political Analysis*, 15, 2007, which I will refer to as PT in the following. The remarks below will respond to the authors' comments published in this issue, which I will refer to as P&T.

PT's Equation (1) and the explanatory text define the model they intend to analyze:

> Recall the data-generating process (DGP) of a FE model with time-invariant variables:
>
> $$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + u_i + \varepsilon_{it}, \tag{1}$$
>
> where the x variables are time varying and the z variables are assumed to be time invariant, u_i denotes the $N-1$ unit-specific effects (FE) of the DGP, is the independent and identically distributed error term, α is the intercept of the base unit, and β and γ are the parameters to be estimated (127).

After a series of steps, PT arrive at the following two paragraphs that conclude their Section 3:

> In stage 3, we rerun the full model without the unit effects but include the unexplained part h_i of the decomposed unit FE vector obtained in stage 2. *This stage is estimated by pooled OLS.* [Emphasis added.]
>
> $$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + \delta h_i + \varepsilon_{it}. \tag{7}$$
>
> By design, h_i is no longer correlated with the vector of the 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, but this bias is of course just the normal omitted variable bias. Yet, given that the estimated unit effects $\hat{u}_i$ consist of much more than the real unit effect u_i and since we cannot disentangle the true elements of u_i from the between variation of the observed and included variables, researchers necessarily face a choice between using as much information as possible and using an unbiased estimator. The fevd procedure thus gives as much power as possible to the available variables unless the within variation is sufficiently large to guarantee efficient estimation.
>
> The estimation of stage 3 proves necessary for various reasons. First of all, *only the third stage allows obtaining the correct SEs*. [Emphasis added.] Not correcting the degrees of freedom leads to a potentially serious underestimation of SEs and overconfidence in the results . . . (129)

(Some additional irrelevant commentary about robust estimation at the end of the second paragraph is omitted.)

The authors recommend a computation used to compute the third step of the FEVD estimator. The second paragraph is the sole (albeit, only implicit) mention in PT about how one should compute SEs for the FEVD estimator. Upon reading this prescription, I asked what the analyst would obtain if they carried out the recommended computation and whether that computation would, indeed, produce the ''correct standard errors,'' as claimed by PT. Pooled ordinary least squares (OLS) at stage 3 produces estimates of (α, β, γ, δ) and what would appear to be a covariance matrix for the full set of estimates. PT state next that ''only the third stage allows obtaining the correct SEs.'' They go on to suggest that the third stage is making a correction for degrees of freedom, but it is not clear what that correction is, as will be evident shortly. PT do not state that the pooled OLS SEs for $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\delta}$ computed at stage 3 should be discarded. They make no further comment about the covariance matrix computed at step 3, so I was left to assume that they meant for pooled OLS estimation of equation (7) to be what one would normally understand pooled OLS to be.

The remainder of my paper explored this ''pooled OLS estimator.'' My first derivation proved that the estimator of β computed at step 3 is identical to the least squares dummy variable (LSDV) estimator at step 1. P&T claim this is obvious. Perhaps so, but that was not stated in PT and, in fact, proving it theoretically, though it can be done with a shorter proof than mine, does require some involved algebra. I also established that the covariance matrix computed for $\hat{\beta}$ at step 3 is smaller than the appropriate matrix that would be computed at step 1. This is not a matter of degrees of freedom. There is hint of an issue with degrees of freedom noted above—the pooled OLS estimator appears to be based on $NT - 1 - K - M - 1$ degrees of freedom. But, this neglects the fact that the dummy variable coefficients have been computed at step 1, so at a minimum, the degrees of freedom should be reduced by $N - 1$. But, this correction does not redeem the matrix computation based on the pooled OLS estimator in equation (7). (Nor, is it necessarily what PT had in mind for ''correcting the degrees of freedom''—that is not made clear in PT.) The matrix, itself, is smaller than its appropriate counterpart, by construction. The appropriate matrix actually appears in PT's equation (8), but that is not the matrix that is produced by pooled OLS estimation of equation (7).[1] The matrix produced by equation (7) is, by construction, always smaller, possibly far smaller than the one in equation (8), as I demonstrated in a numerical example in my paper. It was in this sense that I argued that the efficiency gains are ''illusory''—the FEVD estimator of β at step 3 seems to be more efficient than it really is because of this algebraic result. The standard errors computed for $\hat{\beta}$ at step 3 are too small.

That leaves the FEVD estimator of γ. I also proved that the estimator of γ computed at step 2 is algebraically identical to that at step 3, so there is an ambiguity as to which one should use for the covariance matrix of the estimator of γ, step 2 or step 3. Here, it is unclear what is ''right,'' but it was also clear that the result at step 3 was not ''correct.'' I suggested two possibilities to resolve the ambiguity, one based on a known result that appears in my textbook and others and a second, ''robust'' covariance matrix along the lines of the White estimator.

My final conclusion based on the preceding was that *in the context of their model in equation (1) and under their assumptions that we all agree on*, step 3 of the FEVD estimator was worse than merely a waste of effort. Although it produces the right coefficient estimates, it produces the wrong SEs for the estimator of β and an ambiguous result for the SEs for the estimator of γ. My final recommendation was that researchers who desire to use this technique should not carry out step 3, but, rather should rely entirely on step 1 for β and step 2 plus a side calculation for γ. Along the way, I explored the theoretical implications for identification and estimation of a parameter, γ, on time-invariant variables in a FE model, in the presence of a complete set of ''unit effects'' represented by a set of dummy variables.

In reply to P&T, I note the following: I did not suggest that the FEVD estimator is inconsistent. Not a single word of my paper was about statistical properties of the FEVD estimator, finite sample or asymptotic. My use of the term ''illusory'' was explained above and was entirely appropriate—a practitioner who carries out the recommended computation will be misled at step 3 by the pooled OLS results. The SEs computed for $\hat{\beta}$ at step 3 are indeed ''too small'' in comparison to the long known result that even appears (after the aforementioned correction) in PT's equation (8). P&T have not refuted any of my results.

P&T assert that some of my comments are ''obsolete'' because they updated their *Stata* program in 2010. None of my discussion has anything to do with their *Stata* program (which to this day, I have still never seen). At one point, they also make reference to an ''option,'' apparently a feature of their computer program. But, a fair amount of mechanical detail, including the crucial statement about how to compute SEs, is simply omitted from PT; presumably the reader will find it in the *Stata* program (or will be asked [mistakenly] to trust the program to do the right calculations and not find it at all). The econometric methodology advocated in PT was proposed to the community by publication in this journal. *Stata* is not econometric methodology, and new econometric methods are not established or advanced simply by posting

[1] PT's equation (8) contains an error. The $\hat{\sigma}_u^2$ in that expression should be $\hat{\sigma}_\varepsilon^2$. The error is substantive; $\hat{\sigma}_u^2$ and $\hat{\sigma}_\varepsilon^2$ could differ by orders of magnitude.

code online in a user written .ado file. Practitioners (and theorists) who are not *Stata* users have only the results in *Political Analysis* (and, now, this exchange) to inform them of the relevant theory of this estimator (such as it is). If the claims made for this estimator are to stand up to scrutiny (and reasonable people may doubt they will), and the estimator survives the contest with Hausman and Taylor (H&T) and other candidates, the analysis should be shown to the community of researchers in a more broadly viewed and accepted arena than buried in a *Stata* .ado file.

I have been shown an application of PT's newer version of xtfevd to the data I used in my paper. It does, indeed, faithfully reproduce the LSDV estimator and estimated covariance matrix for β. So, in practical terms, *at least for Stata users of this version of the program*, that part of my paper is indeed moot or merely pedagogical. For the rest of the world that does not use *Stata* (or xtfevd, whatever the current version is), I stand my ground. This leaves several issues on the table as regards estimation of γ. P&T and I still differ on the appropriate covariance matrix to compute for the step 2 estimator, and on my assertion that their step 3 estimator is incorrect. (My opinion that step 3 should not be carried out at all is unchanged.) Then, there are their claims of asymptotic and finite sample efficiency, about which I have thus far had no comment. There are grandiose claims made for the FEVD estimator in both PT and P&T. The statement: ''there cannot be any doubt that FEVD is more efficient than the fixed-effects and HT models (for time invariant, rarely changing and exogenous time-varying variables) and less biased than pooled OLS and random effects (for endogenous time-varying variables in finite samples)'' [P&T, 2–3] is optimistic in the extreme. None of these claims are defended theoretically; they are based on a few Monte Carlo studies. In fact, they are largely speculative. The H&T estimator that figures prominently in this discussion is a straw man. In PT/P&T, there is an explicit assumption that *every TIV is uncorrelated with* u_i. Part of my discussion demonstrated that these M moment equations are essential to identification of γ (and its meaning) in their model, and the entire apparatus must be viewed differently if this assumption is not met. H&T, on the other hand, explicitly assume that *some elements of* x_{it} *and some elements of* z_i *are correlated with* u_i. H&T and PT/P&T are discussing different models. Directly to the point, in H&T's own framework, the first step is LSDV, exactly the same as PT. But, if it were assumed at the outset that z_i and u_i were orthogonal, one would generally not want to follow through with H&T—the availability of the consistent estimator of all of β and the full set of orthogonality conditions for estimation of γ by least squares makes H&T's second to fourth steps moot. It is designed for a different purpose. As noted above, contrary to the grandiose claims made by P&T, there can be doubt.