On the fixed-effects vector decomposition

Breusch, Trevor and Ward, Michael B and Nguyen, Hoa and Kompas, Tom

Australian National University

March 2010

On the Fixed-Effects Vector Decomposition

Trevor Breusch[1,2]
Michael B. Ward
Hoa Nguyen
Tom Kompas

Crawford School of Economics and Government
The Australian National University
Acton, ACT 0200
Australia

[1]The authors are respectively Professor, Associate Professor, Ph.D. candidate, and Professor at the Crawford School of Economics and Government of The Australian National University.

[2]Corresponding author: Trevor.Breusch@anu.edu.au

On the Fixed-Effects Vector Decomposition

ABSTRACT: This paper analyses the properties of the fixed-effects vector decomposition estimator, an emerging and popular technique for estimating time-invariant variables in panel data models with unit 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 which 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 are incorrect. (5) We recommend an alternative shrinkage estimator that has superior risk properties to the decomposition estimator, unless the endogeneity problem is known to be small or no relevant instruments exist.

1. INTRODUCTION

Researchers in many fields seek to exploit the advantages of 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 which might otherwise bias the estimates. However, traditional panel analysis techniques have difficulty when some of the explanatory variables have little or no variation across time within a group. We consider here 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 for such time-invariant or slowly-changing variables.

From the earliest days of panel data methods, there has been a tension between alternative treatments of the unobserved individual effects as either fixed or random. Mundlak (1978) clarified the issues by showing that in the extant literature individual effects were always essentially treated as random, regardless of the estimator used. But, unknown correlation between the unobserved individual effects and the observed individual characteristics would motivate the use of the so-called fixed-effects (FE) estimator even though the underlying model was of random-effects. However, fixed-effects is a blunt instrument for controlling correlation between observed and unobserved characteristics, because it ignores any systematic average differences between individuals. 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 (that is, slowly-changing over the longitudinal dimension) will have little explanatory power, and will result in imprecise coefficient estimates that have large standard errors.

Hausman and Taylor (1981) showed that a better estimator is available if some of the explanatory variables are known to be uncorrelated with the unobserved individual effect, thus described as *exogenous* explanatory variables. The Hausman-Taylor (HT) estimator is an instrumental variables (IV) procedure that combines aspects of both fixed-effects 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 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 150 references in Google Scholar to this emerging estimation methodology. Several empirical studies report standard errors 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 standard errors for coefficients of both the time-invariant and time-varying variables are uniformly *too small.* In the case of the latter variables, the discrepancy in the FEVD standard errors 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 which 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 propose an alternative shrinkage estimator which combines the desirable small sample properties of FEVD with the desirable large sample properties of the HT estimator. Our Monte Carlo evidence shows this shrinkage estimator to have superior risk to either estimator alone over a wide region of the parameter space.

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

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

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 ϵ_{it} are all unobserved. 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 which 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 instrumental variables technique developed by Hausman and Taylor.

[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.

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 nullspace of M. For example,

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

is the matrix which 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 which 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.

The FEVD *Estimator*

The FEVD 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 a fixed effects regression of y on the time-varying X. The moment condition corresponding to a fixed effects 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 $\big(P_D(y - Xb) - Zg\big)' Z = 0$. 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 (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 (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 (7) and the fact that $Q_D P_D = 0$. □

Instrumental Variables 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 fixed effects 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

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

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

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

Using straightforward algebraic manipulation of (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 OLS formula to the Stage 3 regression. Therefore,

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

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

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

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

[2] The usual OLS formula for the standard errors 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_ϵ^2 for large N and small T. Plümper and Troeger (2007a, p. 129) mention briefly an adjustment to the degrees of freedom and, although they do not give an explicit

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

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

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

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

Note that this is exactly the textbook fixed effects variance formula.

Now we note from (16) that s^2 is a consistent estimator of σ_ϵ^2. However, the matrices in the FEVD formula (17) and the correct formula (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 $P_DP_{[Z,h]} = P_{[Z,h]}$. Such a relationship between projection matrices implies that $P_D - P_{[Z,h]}$ is positive semi-definite (in matrix short-hand, $P_D \geq P_{[Z,h]}$). So, $Q_D \leq Q_{[Z,h]}$. That $(X'Q_{[Z,h]}X)^{-1} \leq (X'Q_DX)^{-1}$ follows immediately. This inequality will almost always be strict because the $p+1$ variables

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_ϵ^2, but it is nonstandard and slightly biased. To sharpen the subsequent analysis, we use the standard unbiased estimator of σ_ϵ^2, in which $dof = NT - N - k$.

$[Z, h]$ cannot span the whole of the N-dimensional space of group operator D, and the X's have arbitrary within-group variation. $\square$

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 standard errors 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 (15). Using an alternative representation of the partitioned inverse, this submatrix can be written as

$$V_{\text{FEVD}}(\gamma) = s^2(Z'Z)^{-1}\Big(I + Z'[X,h]\big([X,h]Q_Z[X,h]\big)^{-1}[X,h]'Z(Z'Z)^{-1}\Big). \quad (19)$$

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 (19),

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

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

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

Again, s^2 is a consistent estimator of σ_ϵ^2, so the first term in (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, $(X'Q_DX)^{-1} \geq (X'Q_{[Z,h]}X)^{-1}$ with almost certain strict inequality, so the last term in the FEVD variance formula (20) understates the corresponding term in the correct variance expression (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. $\square$

In general the FEVD variance formula for γ is systematically *biased* and *inconsistent*. The usual reported standard errors 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 standard errors, are unaffected by these parameters. By increasing either of both of these parameters, with everything else held constant, the extent of the downward bias in the FEVD variance formula becomes arbitrarily large.

Reported results from the applied empirical literature align with these theoretical results. For example, Belke and Spies (2008) present results for pooled OLS, FE, FEVD, and HT. The coefficients for the time-varying variables included are the same, by construction as Theorem (1) shows, for FE and FEVD. However, most of the

FEVD reported standard errors are 0.00 to the two reported significant digits, and they never exceed 0.01. In contrast the FE standard errors range from 0.03 to 0.23, with a median value of 0.07. A similar pattern emerges for the OLS and FEVD estimates for time-invariant variables. The coefficients are broadly similar, as would be expected since both methods use the time-invariant variables as instruments. In this case, the reported standard errors for FEVD are again never greater than 0.01. In contrast, the standard error for OLS estimates of these coefficients ranges from 0.00 to 0.23 with a median value of 0.10. One would not expect OLS to be generally less efficient, given the underlying instruments for the procedures. Several other applications reported both FE and FEVD results (*e.g.* Caporale et al., 2009; Mitze, 2009; Krogstrup and Wälti, 2008). 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 standard error 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 instrumental variable estimator. By also expressing FEVD in its instrumental variable representation we are able to develop insights into their comparative properties. Hausman and Taylor showed that the standard fixed effects estimator is equivalent to an IV estimator with instrument set $Q_D X$. To that, they add any exogenous elements of X or of Z as further instruments.[3]

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 which uses the instrument set $[Q_D X, X_1, Z_1]$. In contrast, the FEVD procedure is an IV estimator which uses the instrument set $[Q_D X, Z_1, Z_2]$.

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.[4] 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.[5] 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

[3]Hausman and Taylor describe $P_D X$ as the additional instrument, but this interpretation follows Breusch et al. (1989).

[4]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 fixed effects estimator of β is consistent.

[5]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.

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 et al. (1995).

Under a mean-squared error (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.[6] In contrast, for a small sample with a small endogeneity problem, the FEVD estimator may be preferable. One conventional approach to finding a balance would be to select between the competing estimators based on a specification test (Baltagi et al., 2003). If the test rejects the null hypothesis of no difference between FEVD and HT estimators, then HT would be selected. Otherwise, the FEVD 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 standard errors of the final selected estimator alone may be misleading; however, bootstrap techniques which 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. Several authors have suggested shrinkage approaches based on using a weighted average of two estimators when one estimator is efficient and the other is consistent;

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

see for example Feldstein (1974), Mundlak (1978), Green and Strawderman (1991), or Mittelhammer and Judge (2005). If the bias, variance, and covariance of two estimators are known, it is algebraically straightforward to find the weight which minimizes the MSE of a combined estimator. In particular, suppose one estimator ϕ is unbiased. The other estimator χ is biased, but presumably 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 which 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 the FEVD and HT estimators.[7] Plugging these empirical estimates into (22) produces a practical weighted-average estimator. Kazimi and Brownstone (1999) discuss bootstrap approaches to estimating standard errors for shrinkage-type estimators.

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

[7] Estimates of the variance and covariance terms follow from application of the standard IV formula, as in equation (13). Also, the difference between the FEVD and HT estimators provides an estimate of the bias of FEVD, since HT is asymptotically unbiased.

and HT estimators, we consider a pretest estimator and a shrinkage estimator, both of which can be viewed as weighted averages of FEVD and HT for specific weighting rules. The pretest estimator selects between FEVD and HT based on a 95% critical value of the Durbin-Wu-Hausman specification test for exogeneity of z_3 (see *e.g.* Davidson and MacKinnon, 1993, p. 237). The shrinkage estimator assigns weights according to a first-stage empirical estimate of formula (22).

Plümper and Troeger argued for the superiority of the FEVD procedure over the HT approach based on Monte Carlo evidence. While 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 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 (p. 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 (in footnote 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 (Han and Schmidt, 2001). In contrast, with a valid 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

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

Z is valid and relevant.

The second difference is that our simulations account for sampling variation in the random effect. The Plümper and Troeger code generates the X, Z, and u (random effect) data using the Stata 'corr2data' command. This command enforces the desired moments *exactly* in the sample. Unfortunately, this causes the effective sampling variability of the random effect to be zero. To see the problem, consider the case with no X's. In this case, the FEVD procedure is equivalent to an ordinary regression of y on the Z's. The estimator for γ is then

$$\hat{\gamma} = (Z'Z)^{-1}Z'y = \gamma + (Z'Z)^{-1}Z'\epsilon + (Z'Z)^{-1}Z'u. \tag{23}$$

The final term in (23) is identically the same in each replication of the Monte Carlo experiment. This is because the 'corr2data' command forces both $Z'Z$ and $Z'u$ to be constant across each sample. Consequently, the sampling variance calculated in the Plümper and Troeger Monte Carlo study reflects only the variance of ϵ, and not the variance of the random effect u.

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

$$y_{it} = 1 + 0.5x_1 + 2x_2 - 1.5x_3 - 2.5z_1 + 1.8z_2 + 3z_3 + u_i + \epsilon_{it}. \tag{24}$$

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 ϵ 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 $\text{cov}(z_3, u)$ 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 $\text{cov}(z_3, u)$. Each experiment has 1000 replications, which vary the random components u and ϵ. 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, while $[x_3, z_3]$ are treated as potentially endogenous.

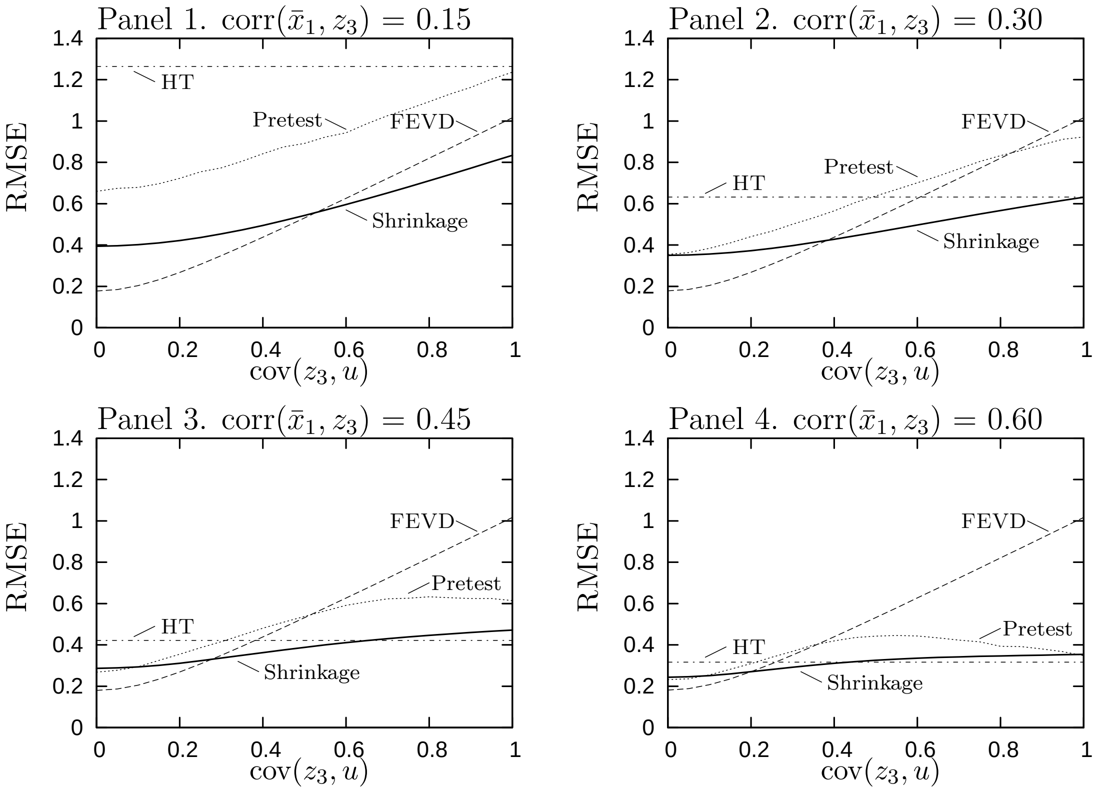
Figure 1 illustrates the simulation results for varying instrument strengths and endogeneity levels. The vertical axis in each panel is the square root of MSE of various estimators for the endogenous time-invariant variable z_3. The horizontal axis of each panel is the covariance between the random effect 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 $\text{corr}(\bar{x}_1, z_3) = 0.15, 0.30, 0.45$, and 0.60 respectively.[11] Note that, within each panel, the HT results are unchanging as a

[9] Conditional on a non-zero 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 corre-

Figure 1: Performance of the four estimators for varying instrument strengths



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

The most notable feature of Figure 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]

lation.

[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.

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 strong priors over endogeneity, using a Bayesian procedure which minimizes risk against that prior might well be the best approach. However, usually, the investigator will be using FE, or HT, or FEVD precisely because of concern that endogeneity might be a significant problem.

Instead of basing the choice of estimator on prior guesses about endogeneity, the investigator can rely on evidence from within the dataset about the degree of endogeneity. 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, and never is much worse than the better of the two. It does this, not by selecting the better of the two basic estimators (as the pretest attempts), but by merging the best qualities of both. The HT estimator offers lower bias, while the FEVD estimator offers lower variance. The shrinkage estimator attempts to make an optimal trade-off over these features in a weighted average of the two basic estimators. Even though the optimal weights must be approximated from the data, the Monte Carlo evidence suggests that the shrinkage estimator would almost certainly be the best choice in the absence of prior information that the endogeneity problem is quite small.

Plümper and Troeger conclude that "the vector decomposition model performs better than the Hausman-Taylor model, pooled OLS, and the [random effects] model." In contrast, we suggest that none of these estimators is likely to be dominant, without strong prior information about endogeneity. Instead, some form of model averaging either through the shrinkage approach discussed here, or through alternatives such as Bayesian model averaging (Hoeting et al., 1999), will generally be the most robust approach.

REFERENCES

Baltagi, B., G. Bresson, and A. Pirotte (2003). Fixed effects, random effects or Hausman–Taylor? A pretest estimator. *Economics Letters 79*(3), 361–369.

Belke, A. and J. Spies (2008). Enlarging the EMU to the east: What effects on trade? *Empirica 35*(4), 369–89.

Bound, J., D. Jaeger, and R. 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*(430), 443–50.

Breusch, T., G. Mizon, and P. Schmidt (1989). Efficient estimation using panel data. *Econometrica 57*(3), 695–700.

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*(2), 189–206.

Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics.* Oxford University Press.

Feldstein, M. (1974). Errors in variables: A consistent estimator with smaller MSE in finite samples. *Journal of the American Statistical Association 69*(348), 990–96.

Green, E. and W. Strawderman (1991). A James-Stein type estimator for combining unbiased and possibly biased estimators. *Journal of the American Statistical Association 86*(416), 1001–06.

Han, C. and P. Schmidt (2001). The asymptotic distribution of the instrumental variable estimators when the instruments are not correlated with the regressors. *Economics Letters 74*(1), 61–66.

Hausman, J. and W. Taylor (1981). Panel data and unobservable individual effects. *Econometrica 49*(6), 1377–98.

Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science 14*(4), 382–401.

James, W. and C. Stein (1961). Estimation with quadratic loss. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, pp. 361–79. University of California Press.

Kazimi, C. and D. Brownstone (1999). Bootstrap confidence bands for shrinkage estimators. *Journal of Econometrics 90*(1), 99–127.

Krogstrup, S. and S. Wälti (2008). Do fiscal rules cause budgetary outcomes? *Public Choice 136*(1), 123–138.

Mittelhammer, R. and G. Judge (2005). Combining estimators to improve structural model estimation and inference under quadratic loss. *Journal of Econometrics 128*(1), 1–29.

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

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*(1), 69–85.

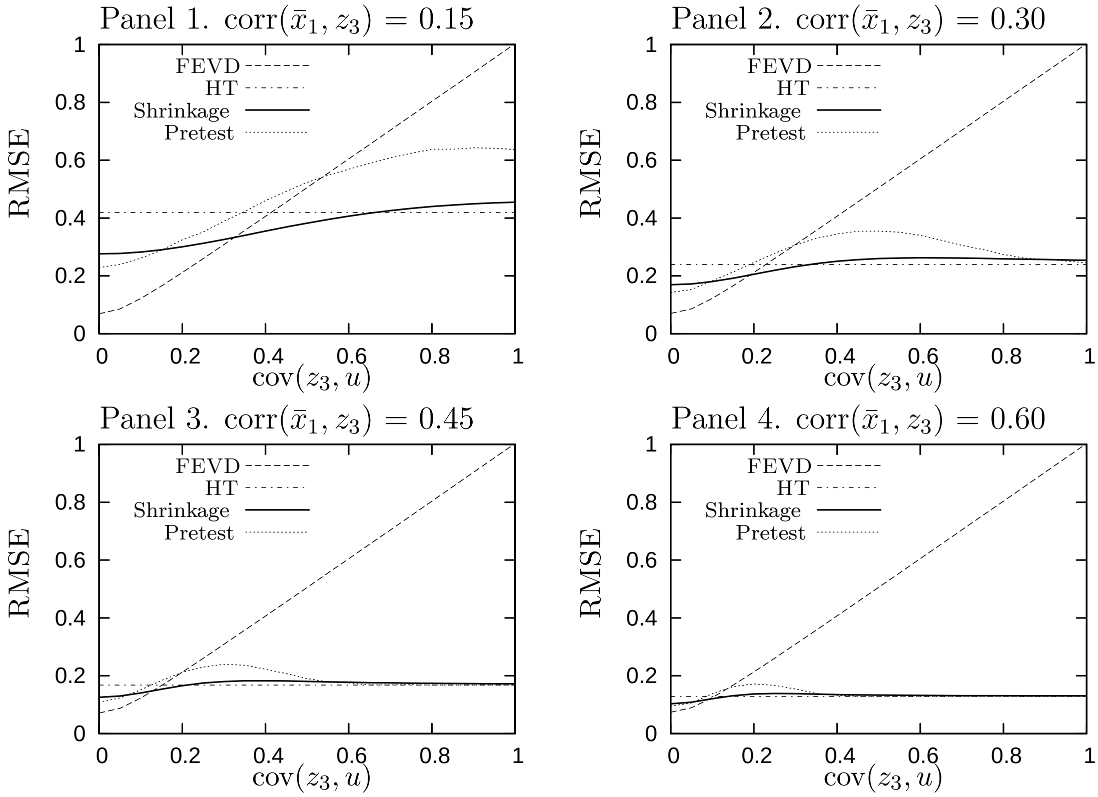
Plümper, T. and V. Troeger (2007a). Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis 15*(2), 124–39.

Plümper, T. and V. Troeger (2007b). xtfevd.ado version 2.00 beta. Accessed from http://www.polsci.org/pluemper/xtfevd.ado.

Wong, K. (1997). Effects on inference of pretesting the exogeneity of a regressor. *Economics Letters 56*(3), 267–71.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

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



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 Figure 2 illustrates the same parameter settings as the corresponding panel in Figure 1. The simulation code for the figures is identical, except for the N and T settings. While 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.