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: $y = x_1 + x_2 + z_1 + z_2 + u + \text{eps}$, $\text{corr}(z_2,u) = 0.5$, $\text{corr}(z_2,x_1) = 0.5$, $\text{corr}(z_2,\text{ext. instr.}) = 0.5$, 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 $\text{cov}(z_3,u)$ set in the experimental design. All other variables are uncorrelated with u, and the variance of u conditional on all variables is 1."

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

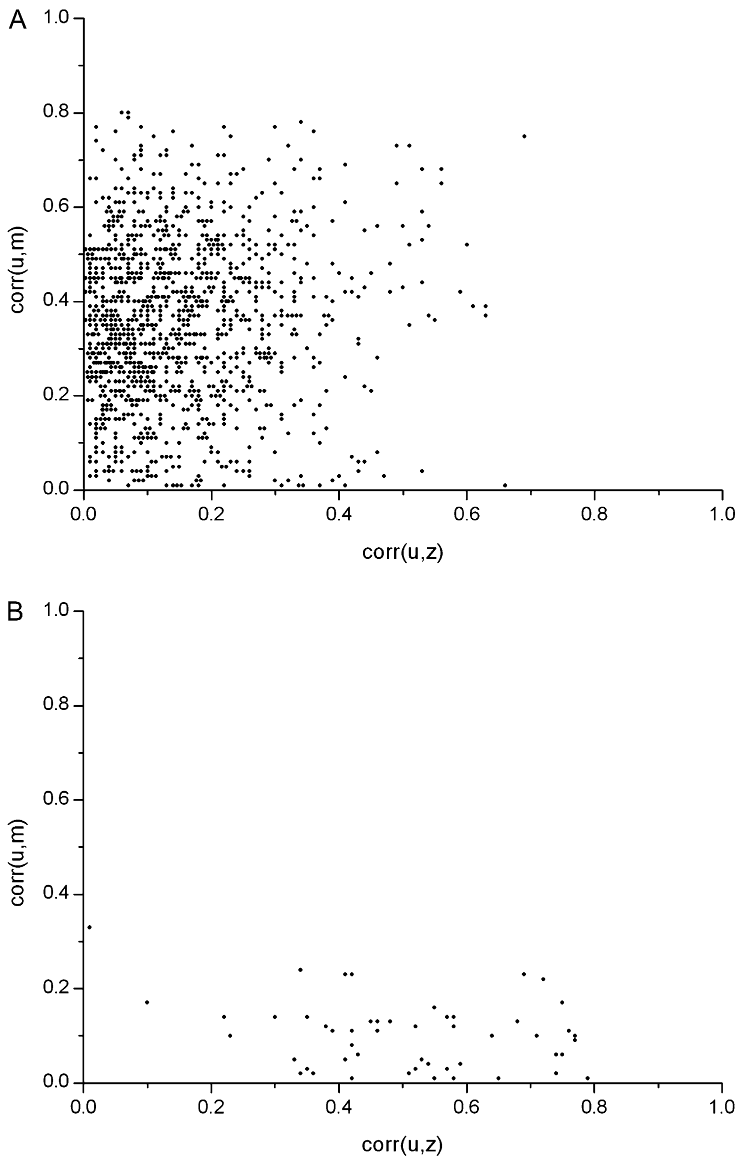

Fig. 1 (a) Constellations in which FEVD significantly outperforms BWNK's shrinkage model ($N = 30$, $T = 20$). (b) Constellations in which BWNK's shrinkage model significantly outperforms FEVD ($N = 30$, $T = 20$).

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

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

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

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.