Advance Access publication February 28, 2007

Political Analysis (2007) 15:124–139
doi:10.1093/pan/mpm002

Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects

Thomas Plümper and Vera E. Troeger

Department of Government, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom, and Max-Planck-Institute of Economics, Kahlaische Strasse 10, 07745 Jena, Germany
e-mail: tpluem@essex.ac.uk (corresponding author)
e-mail: vtroe@essex.ac.uk

This paper suggests a three-stage procedure for the estimation of time-invariant and rarely changing variables in panel data models with unit effects. The first stage of the proposed estimator runs a fixed-effects model to obtain the unit effects, the second stage breaks down the unit effects into a part explained by the time-invariant and/or rarely changing variables and an error term, and the third stage reestimates the first stage by pooled OLS (with or without autocorrelation correction and with or without panel-corrected SEs) including the time-invariant variables plus the error term of stage 2, which then accounts for the unexplained part of the unit effects. We use Monte Carlo simulations to compare the finite sample properties of our estimator to the finite sample properties of competing estimators. In doing so, we demonstrate that our proposed technique provides the most reliable estimates under a wide variety of specifications common to real world data.

1 Introduction

The analysis of "pooled" data has important advantages over pure time-series or cross-sectional estimates—advantages that may easily justify the extra costs of collecting information in both the spatial and the longitudinal dimension. Many applied researchers rank the ability to deal with unobserved heterogeneity across units most prominently. They pool data just for the purpose of controlling for the potentially large number of unmeasured explanatory variables by estimating a "fixed-effects" (FE) model.

Yet, these clear advantages of the FE model come at a certain price. One of its drawbacks, the problem of estimating time-invariant variables in panel data analyses[1] with unit effects, has widely been recognized: since the FE model uses only the within variance for the estimation and disregards the between variance, it does not allow the estimation of

Authors' note: Earlier versions of this paper have been presented at the 21st Polmeth Conference at Stanford University, Palo Alto, July 29–31, 2004, the 2005 MPSA Conference in Chicago, April 7–10, and the APSA Annual Conference 2005 in Washington, September 1–4 2005. We thank the referees of *Political Analysis* and Neal Beck, Greg Wawro, Donald Green, Jay Goodliffe, Rodrigo Alfaro, Rob Franzese, Jörg Breitung, and Patrick Brandt for helpful comments on previous drafts. Any remaining deficiencies are our responsibility.

[1]This article is about time-series–cross-sectional (TSCS) data as defined by Beck and Katz (1995) and Beck (2001). Yet, our procedure can also be applied to panels with short time series. Note that demeaning can be problematic when the number of periods is low.

time-invariant variables (Baltagi 2001; Wooldridge 2002; Hsiao 2003). A second drawback of the FE model (and by far the less recognized one) results from its inefficiency in estimating the effect of variables that have very little within variance. Typical examples in political science include institutions, but political scientists have used numerous variables that show much more variation across units than over time. An inefficient estimation is not merely a nuisance leading to somewhat higher SEs. Inefficiency leads to highly unreliable point estimates and may thus cause wrong inferences in the same way a biased estimator could. Therefore, the inefficiency of the FE model in estimating variables with low within variance needs to be taken seriously.

This article discusses a remedy to the related problems of estimating time-invariant and rarely changing variables in FE models with unit effects. We suggest an alternative estimator that allows estimating time-invariant variables and that is more efficient than the FE model in estimating variables that have very little longitudinal variance. We call this superior alternative "fixed effects vector decomposition" (fevd) model, because the estimator decomposes the unit FE into an unexplained part and a part explained by the time-invariant or the rarely changing variables. The fevd technique involves the following three steps: in the first step, the procedure estimates the unit FE by running a FE estimate of the baseline model. In the second step, the procedure splits the unit effects into an explained and an unexplained part by regressing the unit effects on the time-invariant and/or rarely changing explanatory variables of the original model. Finally, the third stage performs a pooled-OLS estimation of the baseline model by including all explanatory time-variant variables, the time-invariant variables, the rarely changing variables, and the unexplained part of the FE vector. This third stage allows computing correct SEs for the coefficients of the (almost) invariant variables. In addition, one can conveniently use this stage to adjust for serial correlation of errors.[2]

Based on Monte Carlo simulations, we demonstrate that the vector decomposition model has better finite sample properties in estimating models that include either time-invariant or almost time-invariant variables correlated with unit effects than competing estimators. In the analyses dealing with the estimation of time-invariant variables, we compare the vector decomposition model to the FE model, the random effects (RE) model, pooled OLS and the Hausman-Taylor procedure. We find that whereas the FE model does not compute coefficients for the time-invariant variables, the vector decomposition model performs far better than pooled OLS, RE, and the Hausman-Taylor procedure if both time-invariant and time-varying variables are correlated with the unit effects.

The analysis of the rarely changing variables takes these results one step further. Again based on Monte Carlo simulations, we show that the vector decomposition method is more efficient than the FE model and thus gives more reliable estimates than the FE model under a wide variety of constellations. Specifically, we find that the vector decomposition model performs better that the FE model when the ratio between the between variance and the within variance (b/w ratio) is large, when the overall R^2 is low, and when the correlation between the time-invariant or rarely changing variable and the unit effects is low.

In a substantive perspective, this article contributes to an ongoing debate about the pros and cons of FE models (Beck 2001; Beck and Katz 2001; Green, Kim, and Yoon 2001; Plümper, Troeger, and Manow 2005; Wilson and Butler 2007). Although the various

[2]This procedure is superficially similar to that suggested by Hsiao (2003, 52). However, Hsiao only claims that his estimate for time-invariant variables is consistent as N approaches infinity. We are interested in the small sample properties of our estimator and thus explore TSCS data. Hsiao (correctly) notes that his estimate is inconsistent for TSCS. Moreover, he neither provides SEs for his estimate nor compares his estimator to others.

parties in the debate put forward many reasons for and against FE models, this paper analyzes the conditions under which the FE model is inferior to alternative estimation procedures. Most importantly, it suggests a superior alternative for the cases in which the FE model's inefficiency impedes reliable point estimates.

We proceed as follows: in Section 2, we illustrate the estimation problem and discuss how applied researchers dealt with it. In Section 3, we describe the econometrics of the fevd procedure in detail. Section 4 explains the setup of the Monte Carlo experiments. Section 5 analyzes the finite sample properties of the proposed fevd procedure relative to the FE and the RE model, the pooled-OLS estimator, and the Hausman-Taylor procedure in estimating time-invariant variables. Section 6 presents Monte Carlo analyses for rarely changing variables in which we—without loss of generality—compare only the FE model to the vector decomposition model. Section 7 concludes.

2 Estimation of Time-Invariant and Rarely Changing Variables

Time-invariant variables can be subdivided into two broadly defined categories. The first category subsumes variables that are time invariant by definition. Often, these variables measure geography or inheritance. Switzerland and Hungary are both landlocked countries, they are both located in Central Europe, and there is little nature and (hopefully) politics will do about it for the foreseeable future. Along similar lines, a country may or may not have a colonial heritage or a climate prone to tropical diseases.

The second category covers variables that are time invariant for the period under analysis or because of researchers' selection of cases. For instance, constitutions in post-war OECD countries have proven to be highly durable. Switzerland has been a democracy since 1291 and the United States has maintained a presidential system since the adoption of the Constitution. Yet, by increasing the number of periods and/or the number of cases it would be possible to render these variables time-variant.

A small change in the sample can turn a time-invariant variable of the second category into a variable with very low within variation—an almost time-invariant or rarely changing variable. The level of democracy, the status of the president, electoral rules, central bank autonomy, or federalism—to mention just a few—do not change often even in relatively long pooled time-series data sets. Other politically relevant variables, such as the size of the minimum winning coalition, and the number of veto players change more frequently, but the within variance, the variance over time, typically falls short of the between variance, the variance across units. The same may hold true for some macroeconomic aggregates. Indeed, government spending, social welfare, tax rates, pollution levels, or per capita income change from year to year, but panels of these variables can still be dominantly cross-sectional.

Unfortunately, the problem of rarely changing variables in panel data with unit effects remained by-and-large undiscussed.[3] Since the FE model can compute a coefficient if regressors are almost time invariant, it seems fair to say that most applied researchers have accepted the resulting inefficiency of the estimate without paying too much attention to it. Yet, as Nathaniel Beck has unmistakenly formulated: "Although we can estimate [a model] with slowly changing independent variables, the fixed effect will soak up most of the explanatory power of these slowly changing variables. Thus, if a variable ... changes over time, but slowly, the fixed effects will make it hard for such variables to

[3]None of the three main textbooks on panel data analysis (Baltagi 2001; Wooldridge 2002; Hsiao 2003) refers explicitly to the inefficiency of estimating rarely changing variables in a FE approach.

appear either substantively or statistically significant" (Beck 2001, 285). Perhaps even more importantly, inefficiency does not just imply low levels of significance; point estimates are also unreliable since the influence of the error on the estimated coefficients becomes larger as the inefficiency of the estimator increases.

In comparison, by far more attention was devoted to the problem of time-invariant variables. With the FE model not computing coefficients for time-invariant variables, most applied researchers seem to have estimated empirical models that include time-invariant variables by RE models or by pooled OLS (see, e.g., Knack 1993; Huber and Stephens 2001; Acemoglu et al. 2002; Elbadawi and Sambanis 2002). Acemoglu et al. (2002) justify not controlling for unit effects by stating the following: "Recall that our interest is in the historically determined component of institutions (that is more clearly exogenous), hence not in the variations in institutions from year-to-year. As a result, this regression does not (cannot) control for a full set of country dummies." (Acemoglu et al. 2002, 27)

Clearly, both the RE model and pooled OLS are inconsistent and biased when regressors are correlated with the unit effects. Employing these models trades the unbiased estimation of time-varying variables for the ability to compute estimates of time-invariant variables. Thus, they may be a second-best solution if researchers are solely interested in the coefficients of the time-invariant variables.[4]

In contrast, econometric textbooks typically recommend the Hausman-Taylor procedure for panel data with time-invariant variables and correlated unit effects (Hausman and Taylor 1981; see Wooldridge 2002, 325–8; Hsiao 2003, 53). The estimator attempts to overcome the bias of the RE model in the presence of correlated unit effects and the solution is standard: use appropriate instruments for endogenous variables. In brief, this procedure estimates a RE model and uses exogenous time-varying variables as instruments for the endogenous time-varying variables and exogenous time-invariant variables plus the unit means of the exogenous time-varying variables as instruments for the endogenous time-invariant variables (textbook characterizations of the Hausman-Taylor model can be found in Wooldridge [2002, 225–8] and Hsiao [2003, 53ff]). From an econometric perspective, the procedure provides a consistent solution to the potentially severe problem of correlation between unit effects and time-invariant variables. Unfortunately, the procedure can only work well if the instruments are uncorrelated with the errors and the unit effects and highly correlated with the endogenous regressors. Identifying those instruments is a formidable task especially since the unit effects are unobserved (and often unobservable). Nevertheless, the Hausman-Taylor estimator has recently gained in popularity at least among economists (Egger and Pfaffermayr 2004).

3 Fixed Effects Vector Decomposition

Recall the data-generating process (DGP) of a FE model with time-invariant variables:

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

where the x variables are time-varying and the z variables are assumed to be time invariant,[5] u_i denotes the $N-1$ unit-specific effects (FE) of the DGP, ε_{it} is the independent

[4]The RE model is unbiased only when the pooled-OLS model is unbiased as well. However, the RE model is, under broad conditions, more efficient than the pooled-OLS model.

[5]In Section 5, we assume that one z variable is rarely changing and thus almost time invariant.

and identically distributed error term, α is the intercept of the base unit, and β and γ are the parameters to be estimated.

In the first stage, the fevd procedure estimates a standard FE model. The FE transformation can be obtained by first averaging equation (1) over T:

$$\bar{y}_i = \alpha + \sum_{k=1}^{K} \beta_k \bar{x}_{ki} + \sum_{m=1}^{M} \gamma_m z_{mi} + \bar{e}_i + u_i, \tag{2}$$

where

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}, \quad \bar{e}_i = \frac{1}{T}\sum_{t=1}^{T} e_{it}$$

and e stands for the residual of the estimated model. Then equation (2) is subtracted from equation (1). This transformation removes the individual effects u_i and the time-invariant variables z. We get

$$y_{it} - \bar{y}_i = \beta_k \sum_{k=1}^{K} (x_{kit} - \bar{x}_{ki}) + \gamma_m \sum_{m=1}^{M} (z_{mi} - z_{mi}) + (e_{it} - \bar{e}_i) + (u_i - u_i)$$

$$\equiv \ddot{y}_{it} = \beta_k \sum_{k=1}^{K} \ddot{x}_{kit} + \ddot{e}_{it}, \tag{3}$$

with $\ddot{y}_{it} = y_{it} - \bar{y}_i$, $\ddot{x}_{kit} = x_{kit} - \bar{x}_{ki}$, and $\ddot{e}_{it} = e_{it} - \bar{e}_i$ denoting the demeaned variables of the within transformation. We run this FE model with the sole intention to obtain estimates of the unit effects $\hat{u}_i$. At this point, it is important to note that the "estimated unit effects" $\hat{u}_i$ do not equal the unit effects u_i in the DGP.[6] Rather, these estimated unit effects include all time-invariant variables, the overall constant term, and the mean effects of the time-varying variables x—or, in other words,

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

where β_k^{FE} is the pooled-OLS estimate of the demeaned model in equation (3).

This $\hat{u}_i$ includes the unobserved unit-specific effects as well as the observed unit-specific effects z, the unit means of the residuals $\bar{e}_i$, and the time-varying variables $\bar{x}_{ki}$, whereas u_i only accounts for unobservable unit-specific effects. In stage 2, we regress the unit effects $\hat{u}_i$ from stage 1 on the observed time-invariant and rarely changing variables—the z variables (see equation (5)) to obtain the unexplained part h_i (the residual from regressing the unit-specific effect on the z variables). In other words, we decompose the estimated unit effects into two parts, an explained and an unexplained part that we dub h_i:

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

[6]We follow standard practice by this notation. However, from equation (4) it follows that the FE estimate of the unit effects propels much more to the estimated unit effects. To avoid confusion and maintain consistence with standard textbooks, we stick to this notation—needless to say that it does not make much sense.

The unexplained part h_i is obtained by computing the residuals from equation (5):

$$h_i = \hat{u}_i - \sum_{m=1}^{M} \gamma_m z_{mi}. \tag{6}$$

As we said above, this crucial stage decomposes the unit effects into an unexplained part and a part explained by the time-invariant variables. We are solely interested in the *unexplained* part h_i.

In stage 3, we rerun the full model without the unit effects but include the unexplained part h_i of the decomposed unit FE vector obtained in stage 2. This stage is estimated by pooled OLS.

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

By design, h_i is no longer correlated with the vector of the z variables. If the time-invariant variables are assumed to be orthogonal to the unobserved unit effects—i.e., if the assumption underlying our estimator is correct—the estimator is consistent. If this assumption is violated, the estimated coefficients for the time-invariant variables are biased,[7] but this bias is of course just the normal omitted variable bias. Yet, given that the estimated unit effects $\hat{u}_i$ consist of much more than the real unit effect u_i and since we cannot disentangle the true elements of u_i from the between variation of the observed and included variables, researchers necessarily face a choice between using as much information as possible and using an unbiased estimator. The fevd procedure thus gives as much power as possible to the available variables unless the within variation is sufficiently large to guarantee efficient estimation.

The estimation of stage 3 proves necessary for various reasons. First of all, only the third stage allows obtaining the correct SEs. Not correcting the degrees of freedom leads to a potentially serious underestimation of SEs and overconfidence in the results. Second, the third stage also allows researchers to explicitly deal with the dynamics of the time-invariant variables. This is important since estimating the model requires that heteroscedasticity and serial correlation must be eliminated. If the structure of the data at hand is as such, we suggest running a robust Sandwich estimator or a model with panel-corrected SEs (in stages 1 and 3) and inclusion of the lagged dependent variable (Beck and Katz 1995) and/or modeling the dynamics by Prais-Winsten transformation of the original data in stages 1 and 3.

4 Design of the Monte Carlo Simulations

In what follows we conduct a series of Monte Carlo analyses to compare the finite sample properties of estimators that have frequently been applied to conditions similar to those the fevd estimator is made for or that have been suggested by econometrics textbooks. In particular, we are interested in the finite sample properties of competing estimators.

We follow the literature in using the root mean squared error (RMSE) as criterion. RMSEs provide a unified view of the two main sources of wrong point estimates: bias and inefficiency. King, Keohane, and Verba (1994, 74) highlight the fundamental trade-off

[7] Note that the estimated coefficients of the time-varying variables remain unbiased even in the presence of correlated unit effects. However, the assumptions underlying a FE model must be satisfied (no correlated time-varying variables may exist).

between bias and efficiency: "We would ... be willing to sacrifice unbiasedness ... in order to obtain a significantly more efficient estimator. ... The idea is to choose the estimator with the minimum mean squared error since it shows precisely how an estimator with some bias can be preferred if it has a smaller variance." This potential trade-off between efficiency and unbiasedness implies that the choice of the best estimator typically depends on the sample size. If researchers always went for the estimator with the best asymptotic properties (as typically recommended in econometrics textbooks) they would always choose the best estimator for infinitely large samples. Unfortunately, this estimator could perform poorly in estimating the finite sample at hand.

All experiments use simulated data, which are generated to discriminate between the various estimators, and at the same time exhibit some properties of time-series–cross-sectional and cross-sectional–time-series data. Specifically, the DGP underlying our simulations is as follows:

$$y_{it} = \alpha + \beta_1 x_{1_{it}} + \beta_2 x_{2_{it}} + \beta_3 x_{3_{it}} + \beta_4 z_{1_i} + \beta_5 z_{2_i} + \beta_6 z_{3_i} + u_i + \varepsilon_{it},$$

where the x variables are time-varying and the z variables are time invariant. Both groups are drawn from a normal distribution. u_i denotes the unit-specific unobserved effects and also follows a normal distribution. The idiosyncratic error ε_{it} is white noise and drawn from a standard normal distribution for each run. The R^2 is fixed at 50% for all simulations. x_3 is a time-varying variable correlated with the unit effects u_i, whereas z_3 is time invariant in Section 5 and rarely changing in Section 6. In both cases, z_3 is correlated with u_i. We hold the coefficients of the true model constant throughout all experiments at the following values:

$$\alpha = 1,\ \beta_1 = 0.5,\ \beta_2 = 2,\ \beta_3 = -1.5,\ \beta_4 = -2.5,\ \beta_5 = 1.8,\ \beta_6 = 3.$$

Among these six variables, only variables x_3 and z_3 are of analytical interest since only these two variables are correlated with the unit-specific effects u_i. Variables x_1, x_2, z_1, and z_2 do not covary with u_i. However, we include these additional time-variant and time-invariant variables into the DGP, because we want to ensure that the Hausman-Taylor instrumental estimation is at least just identified or even overidentified (Hausman and Taylor 1981). For the same reason, we let z_1 and z_2 be uncorrelated with the unit effects. Although this assumption seems unrealistic, it is necessary to satisfy the minimum conditions for instruments. This unrealistic assumption thus ensures that the advantages of the fevd estimator over the Hausman-Taylor model cannot be explained by the poor quality of instruments.

We hold this outline of the simulations constant in Section 6, where we analyze the properties of the FE model and the vector decomposition technique in the presence of rarely changing variables correlated with the unit effects. Although the inclusion of the uncorrelated variables x_1, x_2, z_1, and z_2 appears not necessary in Section 6, these variables do not adversely affect the simulations and we keep them to maintain comparability across all experiments.[8] In the experiments, we varied the correlation between x_3 and the unit effects $\text{corr}(x_3, u_i) = \{0.0, 0.1, 0.2, \ldots, 0.9, 0.99\}$ and the correlation between z_3 and the unit effects $\text{corr}(z_3, u_i) = \{0.0, 0.1, 0.2, \ldots, 0.9, 0.99\}$.[9]

[8] z_3 in section 5 is rarely changing, the between and within SD for this variable are changed according to the specifications in Figs. 2–4.

[9] We also varied the number of units (N = 15, 30, 50, 70, 100) and the number of time periods (T = 20, 40, 70, 100). We report these results only in the online appendix. The number of possible permutations of these settings is 2000 that would have led to 2000 times the aggregated number of estimators used in both experiments times 1000 single estimations in the Monte Carlo analyses. In total, this would have given 18 million regressions. However, without loss of generality, we simplified the Monte Carlos and estimated "only" 980,000 single regression models.

Table 1 Average RMSE and bias over 10 permutations times 1000 estimations

	Pooled OLS		*fevd*		*Hausman-Taylor*		*RE*		*FE*	
	RMSE	*Average bias*	*RMSE*	*Average bias*	*RMSE*	*Average bias*	*RMSE*	*Average bias*	*RMSE*	*Average bias*
Time-varying variable x_3	0.187	−0.167	0.103	0.001	0.105	−0.003	0.173	−0.149	0.103	−0.001
Time-invariant variable z_3	0.494	−0.470	0.523	−0.548	1.485	−1.128	0.506	−0.481		
Settings of the parameter held constant										
$N = 30$, $T = 20$										
$\text{corr}(u, x_1) = \text{corr}(u, x_2) = \text{corr}(u, z_1) = \text{corr}(u, z_2) = 0$										
$\text{corr}(u, x_3) = 0.5$										
Settings of the varying parameter										
$\text{corr}(u, z_3) = \{0.1, 0.2, \ldots, 0.9, 0.99\}$										

5 The Estimation of Time-Invariant Variables

We report the RMSE and the bias of the five estimators, averaged over 10 experiments with varying correlation between z_3 and u_i. The Monte Carlo analysis underlying Table 1 holds the sample size and the correlation between x_3 and u_i constant. In other words, we vary only the correlation between the correlated time-invariant variable z_3 and the unit effects $\text{corr}(u, z_3)$.

Observe first that (in this and all following tables) we highlight all estimation results if the estimator performs best or if its RMSE exceeds that of the best estimator by less than 10%. Table 1 reveals that estimators vary widely in respect to the correlated explanatory variables x_3 and z_3. Whereas the vector decomposition model, Hausman-Taylor, and the FE model estimate the coefficient of the correlated time-varying variable (x_3) with almost identical accuracy, pooled OLS, the vector decomposition model, and the RE model perform more or less equally well in estimating the effects of the correlated time-invariant variable (z_3). In sum, only the fevd model performs well with respect to both variables correlated with the unit effects x_3 and z_3.

The poor performance of Hausman-Taylor results from the inefficiency of instrumental variable models. Although it holds true that one can reduce the inefficiency of the Hausman-Taylor procedure by improving the quality of the instruments,[10] all carefully selected instruments have to satisfy two conditions simultaneously: they have to be uncorrelated with the unit effects and correlated with the endogenous variables. Needless to say, finding instruments that simultaneously satisfy these two conditions is a difficult task—especially since the unit effects cannot be observed, but only be estimated.

Pooled OLS and the RE model fail to adequately account for the correlation between the unit effects and both the time-invariant and the time-varying variables. Hence, parameter estimates for all variables correlated with the unit effects are biased. When applied researchers are theoretically interested in both time-varying and time-invariant variables, the fevd technique is superior to its alternatives.

[10]This has been suggested by Amemiya and MaCurdy (1986), Breusch, Mizon, and Schmidt (1989), Baltagi and Khanti-Akom (1990), Baltagi, Bresson, and Pirotte (2003), and Oaxaca and Geisler (2003).

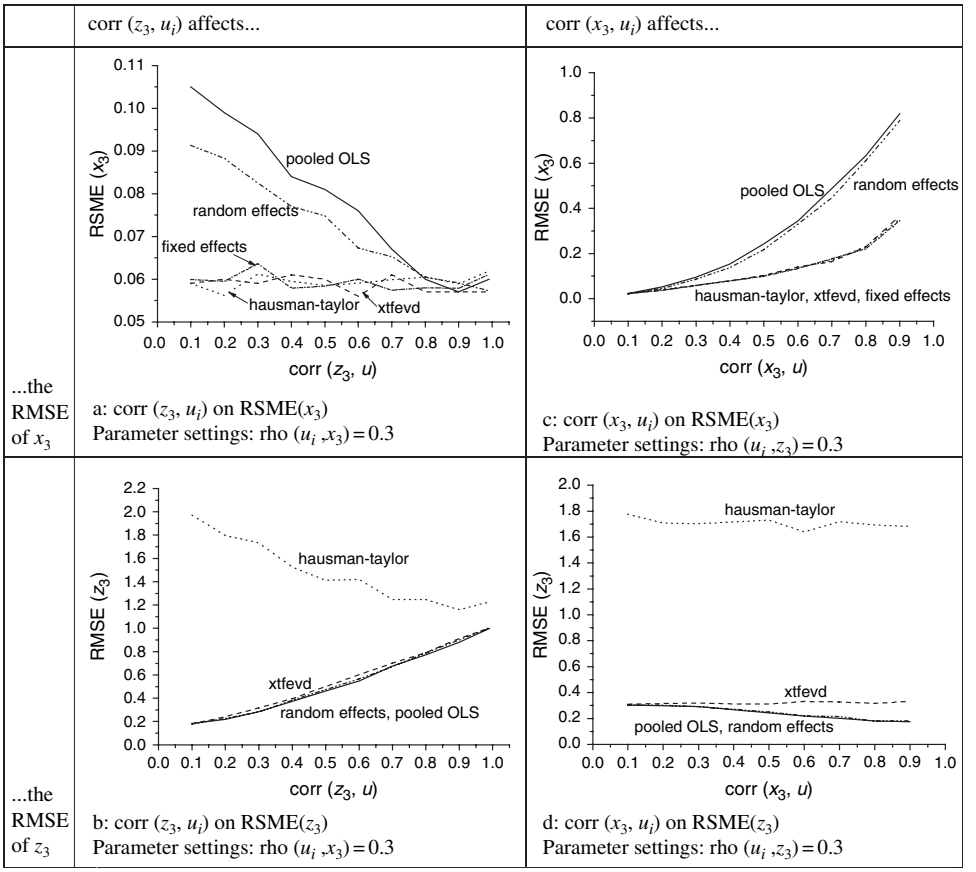


Fig. 1 Change in the RMSE over variation in the correlation between the unit effects and z_3, x_3, respectively; $N = 30$, $T = 20$.

Figures 1a–d allow an equally easy comparison of the five competing estimators. Note, in the simulations underlying these figures, we held all parameters constant and varied only the correlation between the time-invariant variable z_3 and u_i (Fig. 1a and 1b) and the time-varying variable x_3 and u_i (Fig. 1c and 1d), respectively. Figures 1a and 1c display the effect of this variation on the RMSE of the estimates for the time-varying variable x_3 and Fig. 1b and 1d the effect on the coefficient of the time-invariant variable z_3.

Figures 1a–d reestablish the results of Table 1. We find that fevd, RE, and pooled OLS perform equally well in estimating the coefficient of the correlated time-invariant variable z_3, whereas FE, Hausman-Taylor, and fevd are superior in estimating the coefficient of time-varying variable x_3. We find that the advantages of the vector decomposition procedure over its alternatives do not depend on the size of the correlation between the regressors and the unit effects but rather hold over the entire bandwidth of correlations.

Of the models tested here, only the fevd model gives reliable finite sample estimates if the data set to be estimated includes time-varying and time-invariant variables correlated with the unit effects.[11] In the next section, we further explore the fevd estimator and turn to

[11]The online appendix (see the *Political Analysis* Web page for online appendices) demonstrates that this result also holds true when we vary the sample size. The fevd model performs best even with a comparably large T and N.

the perhaps more important issue of how to handle variables that are not time invariant, but whose within variation is low—absolute and relative to the between variation.

6 Estimation of Variables with Low within Variation

One important advantage of the fevd procedure over the Hausman-Taylor procedure and the Hsiao suggestions is that it extends nicely to almost time-invariant variables. Estimation of these variables by FE gives a coefficient, but the estimation suffers from inefficiency that renders the estimated coefficients unreliable (Beck and Katz 2001; Green, Kim, and Yoon 2001). However, if we do not estimate the model by FE, then estimated coefficients are biased if the regressor is correlated with the unit effects. Nevertheless, since it seems not unreasonable to assume that in comparative politics the unit effects are made up primarily of geographical, cultural, and various institutional variables and since most of these variables can in principle be observed, it is not unreasonable to perform an orthogonal decomposition of the explained part and an unexplained part as described above. Clearly, the orthogonality assumption is often incorrect and this will inevitably bias the estimated coefficients of the almost time-invariant variables. However, as we will demonstrate in this section, this bias does under distinguishable conditions less harm than the inefficiency caused by FE estimation. Obviously the performance of fevd will depend on the exact DGP. In our simulations we show that unless the DGP is highly unfavorable for fevd, our procedure performs reasonably well and is generally better than its alternatives.

Before we report the results of the Monte Carlo simulations, let us briefly explain why the estimation of almost time-invariant variables by the standard FE model is problematic due to inefficiency and what that inefficiency does to the estimate. The inefficiency of the FE model results from the fact that it disregards the between variation. Thus, the FE model does not take all the available information into account. In technical terms, the estimation problem stems from the asymptotic variance of the FE estimator:

$$A\,\widehat{\mathrm{var}}(\hat{\beta}^{\mathrm{FE}}) = \hat{\sigma}_u^2 \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{X}_i \right)^{-1}. \tag{8}$$

When the FE model performs the within transformation on a variable with little within variance, the variance of the estimates can approach infinity. Thus, if the within variation becomes very small, the point estimates of the FE estimator become unreliable. In this situation, the FE model does not only compute large SEs, but in addition the sampling variance gets large and therefore the reliability of point predictions is low and the probability that the estimated coefficient deviates largely from the true coefficient increases (see Beck and Katz 2001).

Our Monte Carlo simulations seek to identify the conditions under which the fevd model computes more reliable coefficients than the FE model. Table 2 reports the output of a typical simulation analogous to Table 1.[12]

Results displayed in Table 2 mirror those reported in Table 1. As before, we find that only the fevd procedure gives sufficiently reliable estimates for both the correlated

[12]We also compared the vector decomposition and the FE model to pooled-OLS and the RE model. Since all findings for time-invariant variables carry over to rarely changing variables, indicating that the vector decomposition model dominates pooled-OLS and RE models, we report the results of the RE and pooled-OLS Monte Carlos only in the online appendix.

Table 2 Average RMSE and bias over 10 permutations times 1000 estimations

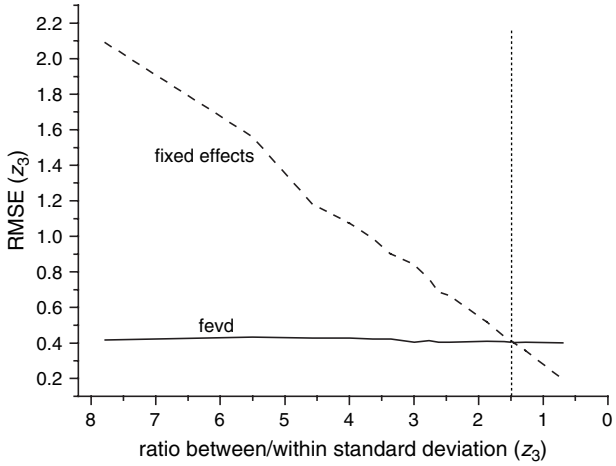
	fevd		*FE*	
	RMSE	*Average bias*	*RMSE*	*Average bias*
Time-varying variable x_3	0.069	0.001	0.069	0.000
Rarely changing variable z_3	0.131	0.001	0.858	0.008
Parameters held constant				
$N = 30$, $T = 20$				
$\text{corr}(u, x_1) = \text{corr}(u, x_2) = \text{corr}(u, z_1) = \text{corr}(u, z_2) = 0$				
$\text{corr}(u, z_3) = 0.3$				
$\text{corr}(u, x_3) = 0.5$				
Between SD $(z_3) = 1.2$				
Varied parameters				
Within SD $(z_3) = \{0.04, \ldots, 0.94\}$				

time-varying x_3 and the rarely changing variable z_3. As expected, the FE model provides far less reliable estimates of the coefficients of rarely changing variables. The fevd model can improve the reliability of the estimation in the presence of variables with low within and relatively high between variance. We also find that pooled OLS and the RE model estimate rarely changing variables with more or less the same degree of reliability as the fevd model but are far worse in estimating the coefficients of time-varying variables. Note that these results are robust regardless of sample size.[13]

Since any further discussion of these issues would be redundant, we do not further consider the RE and the pooled-OLS model in this section. Rather, this section provides answers to two interrelated questions: first, can the vector decomposition model give better estimates (a lower RMSE) than the FE model and second, in case we can answer the first question positively, what are the conditions that determine the relative performance of both estimators? To answer these questions, we assess the finite sample properties of the competing models in estimating rarely changing variables by a second series of Monte Carlo experiments. With one notable exception, the DGP in this section are identical to the one used in Section 5. The exception is that now z_3 is not time invariant but a "rarely changing variable" with a low within variation and a defined between to within variance ratio (b/w ratio).

The easiest way to explore the relative performance of the FE model and the vector decomposition model is to change the ratio between the cross-sectional (between) variance and the time-series (within) variance across experiments. We compute this b/w ratio by dividing the between SD of a variable by the within SD of the same variable. There are two ways to vary this ratio systematically: we can hold the between variation constant and vary the within variation or we can hold the within variation constant and vary the between variation. We use both techniques. In Fig. 2, we hold the between SD constant at 1.2 and change the within SD successively from 0.15 to 1.73, so that the b/w ratio varies between 8 and 0.7. In Fig. 3, we hold the within variance constant and change the between variance.

[13]We reran all Monte Carlo experiments on rarely changing variables for different sample sizes. Specifically, we analyzed all permutations of $N = \{15, 30, 50, 70, 100\}$ and $T = \{20, 40, 70, 100\}$. The results are shown in Table A2 of Appendix A. All findings for rarely changing variables remain valid for larger and smaller samples, as well as for N exceeding T and T exceeding N.



Parameter settings: $N = 30$; $T = 20$; rho(u,x_3) = 0; rho(u,z_3) = 0.3; between SD (z_3): 1.2; within SD (z_3): 0.15…1.73

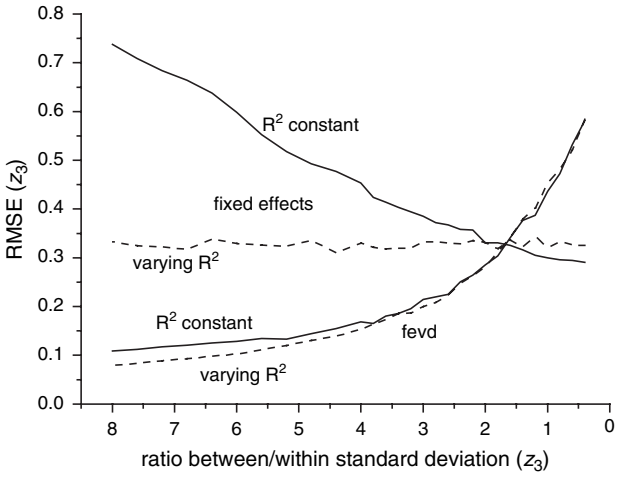
Fig. 2 The ratio of between- to within SD (z_3) on RSME (z_3).

Recall that the estimator with the lower RMSE gives more reliable and better estimates. Hence, Fig. 2 shows that when the within variance increases relative to the between variance, the FE model becomes increasingly reliable. Since the reliability of the vector decomposition model does not change, we find that the choice of an estimator is contingent. If the b/w ratio is smaller than approximately 1.7, the FE estimator performs better than the vector decomposition model. However, the threshold depends on the correlation between the rarely changing variables and the true unit effects. Above this threshold, trading unbiasedness for the efficiency of the vector decomposition model improves the estimates.

We obtain similar results when we change the within variation and keep the between variation constant. Figure 3 shows simultaneously the results of two slightly different experiments. In one experiment (dotted lines), we varied the within variation and kept the between variation and the error constant. In the other experiment, we kept the between variation constant but varied the within variation and the error variance in a way that the fevd R^2 remained constant.

In both experiments, we find the threshold to be at approximately 1.7 for a correlation of z_3 and u_i of 0.3. We can conclude that the result is not merely the result of the way in which we computed variation in the b/w ratio, since the threshold level remained constant over the two experiments.

Unfortunately, the relative performance of the FE model and the vector decomposition model does not solely depend on the b/w ratio. Rather, we also expected, and found, a strong influence of the correlation between the rarely changing variable and the unit effects. The influence of the correlation between the unit effects and the rarely changing variable obviously results from the fact that it affects the bias of the vector decomposition model but does not influence the inefficiency of the FE model. Thus, a larger correlation between the unit effects and the rarely changing variable renders the vector decomposition model worse relative to the FE model. Unfortunately, this correlation is unobservable and an indirect, Hausman-type, test does not (yet) exist. We illustrate the strength of this effect



Parameter settings: $N = 30$; $T = 20$; rho(u,x_3) = 0; rho(u,z_3) = 0.3
between SD (z_3): 0.4…8; within SD (z_3): 1

Fig. 3 The ratio of between- to within SD (z_3) on RSME (z_3).

by relating it to the level of the b/w ratio, at which the FE model and the fevd model give identical RMSE. Accordingly, Fig. 4 displays the dependence of the threshold level of the b/w ratio on the correlation between the rarely changing variable and the unit effects.

Note that, as expected, the threshold b/w ratio is strictly increasing in the correlation between the rarely changing variable and the unobserved unit effects. In the case where the rarely changing variable shows no correlation with u_i, the threshold of the b/w ratio is as small as 0.2. At a correlation of 0.3, fevd is superior to the FE model if the b/w ratio is larger than approximately 1.7; at a correlation of 0.5 the threshold increases to about 2.8, and at a correlation of 0.8 the threshold gets close to 3.8. Therefore, we cannot offer a simple rule of thumb that informs applied researchers of when the estimation of a particular model by fevd gives better results. Perhaps even worse, the correlation between the unit effects and the rarely changing variable cannot be directly observed, because the unit effects are unobservable. However, the odds are that at a b/w ratio of at least 2.8, the variable is better included into the stage 2 estimation of fevd than estimated by a standard FE model.

Applied researchers can improve estimates created by the vector decomposition model by reducing the potential for correlation. To do so, stage 2 of the fevd model needs to be carefully studied. We can reduce the potential for bias of the estimation by including additional time-invariant or rarely changing variables into stage 2. This may reduce bias but is likely to also reduce efficiency. Alternatively, applied researchers can use variables that are uncorrelated with the unit effects as instruments for potentially correlated time-invariant or rarely changing variables—a strategy that resembles the Hausman-Taylor model. Yet, as we have repeatedly pointed out: it is impossible to tell good from bad instruments since the unit effects cannot be observed.

The decision whether to treat a variable as time invariant or varying depends on the b/w ratio of this variable and on the correlation between the unit effects and the rarely changing variables. In this respect, the estimation of time-invariant variables is just a special case of the estimation of rarely changing variables—a special case in which the b/w ratio equals infinity.

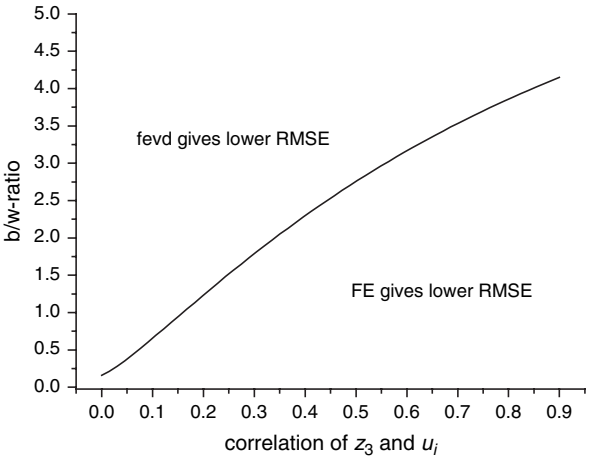


Fig. 4 The correlation between z_3 and u_i and the minimum ratio between the between- and within SD that renders fevd superior to the FE model.

These findings suggest that—strictly speaking—the level of within variation does not influence the relative performance of fevd and FE models. However, with a relatively large within variance, the problem of inefficiency does not matter much—the RMSE of the FE estimator will be low. Still, if the within variance is large but the between variance is much larger, the vector decomposition model will perform better on average. With a large within variance, the actual absolute advantage in reliability of the fevd estimator will be tiny.

From a more general perspective, the main result of this section is that the choice between the FE model and the fevd estimator depends on the relative efficiency of the estimators and on the bias. As King, Keohane, and Verba (1994, 74) have argued, applied researchers should not base their choice of the estimator solely on unbiasedness. Point predictions become more reliable when researchers use the more efficient estimator. The fevd model is more efficient than the FE model since it uses more information. Rather than just relying on the within variance, our estimator also uses the between variance to compute coefficients. But it is biased.

7 Conclusion

Under specific conditions, the vector decomposition model produces more reliable estimates for time-invariant and rarely changing variables in panel data with unit effects than any alternative estimator of which we are aware of. The case for the vector decomposition model is clear when researchers are interested in time-invariant variables. Whereas the FE model does not compute coefficients of time-invariant variables, the vector decomposition model performs better than the Hausman-Taylor model, pooled OLS, and the RE model.

The case for the vector decomposition model is less straightforward when at least one regressor is not strictly time invariant, but shows some variation across time. Nevertheless, under many conditions the vector decomposition technique produces more reliable estimates. These conditions are: first, and most importantly, the between variation needs to be

larger than the within variation and second, the higher the correlation between the rarely changing variable and the unit effects, the worse the vector decomposition model performs relative to the FE model and the higher the b/w ratio needs to be to render fevd more reliable.

From our Monte Carlo results, we can derive the following rules that may inform the applied researcher's selection of an estimator on a more general level: estimation by pooled-OLS or RE models is only appropriate if unit effects do not exist or if the Hausman test suggests that existing unit effects are uncorrelated with the regressors. If either of these conditions is violated, the FE model and the vector decomposition model compute more reliable estimates for time-varying variables. Among these two models, the FE model performs best if the within variance of all regressors of interest is sufficiently large in comparison to their between variance. Otherwise, the efficiency of the fevd model becomes more important than the unbiasedness of the FE model.

References

Acemoglu, Daron, Simon Johnson, James Robinson, and Yunyong Thaicharoen. 2002. Institutional causes, macroeconomic symptoms: Volatility, crises and growth. NBER working paper 9124.

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

Baltagi, Badi H. 2001. *Econometric analysis of panel data.* Chichester, UK: Wiley and Sons.

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

Baltagi, Badi H., and Sophon Khanti-Akom. 1990. On efficient estimation with panel data: An empirical comparison of instrumental variable estimators. *Journal of Applied Econometrics* 5:401–6.

Beck, Nathaniel. 2001. Time-series-cross-section data: What have we learned in the past few years? *Annual Review of Political Science* 4:271–93.

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.

———. 2001. Throwing out the baby with the bath water: A comment on Green, Kim, and Yoon. *International Organization* 55:487–95.

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

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

Egger, Peter, and Michael Pfaffermayr. 2004. Distance, trade and FDI: A Hausman-Taylor SUR approach. *Journal of Applied Econometrics* 19:227–46.

Elbadawi, Ibrahim, and Nicholas Sambanis. 2002. How much war will we see? Explaining the prevalence of civil war. *Journal of Conflict Resolution* 46:307–34.

Green Donald, P., Soo Yeon Kim, and David H. Yoon. 2001. Dirty pool. *International Organization* 55: 441–68.

Greenhalgh, C., M. Longland, and D. Bosworth. 2001. Technological activity and employment in a panel of UK firms. *Scottish Journal of Political Economy* 48:260–82.

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.

Hsiao, Cheng. 1987. Identification. In *Econometrics*, ed. J. Eatwell, M. Milgate, and P. Newman, 95–100. London: W.W. Norton.

———. 2003. *Analysis of panel data.* Cambridge: Cambridge University Press.

Huber, Evelyne, and John D. Stephens. 2001. *Development and crisis of the welfare state. Parties and policies in global markets.* Chicago, IL: University of Chicago Press.

Iversen, Torben, and Thomas Cusack. 2000. The causes of welfare state expansion. Deindustrialization of globalization. *World Politics* 52:313–49.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing social inquiry: Scientific inference in qualitative research.* Princeton, NJ: Princeton University Press.

Knack, Stephen. 1993. The voter participation effects of selecting jurors from registration lists. *Journal of Law and Economics* 36:99–114.

Oaxaca, Ronald L., and Iris Geisler. 2003. Fixed effects models with time-invariant variables. A theoretical note. *Economics Letters* 80:373–7.

Plümper, Thomas, Vera E. Troeger, and Philip Manow. 2005. Panel data analysis in comparative politics. Linking method to theory. *European Journal of Political Research* 44:327–54.

Wilson, Sven E., and Daniel M. Butler. 2007. A lot more to do: The sensitivity of time-series cross-section analyses to simple alternative specifications. *Political Analysis* 10.1093/pan/mpl012.

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