Estimation of Time-invariant Effects in Static Panel Data Models*

M. Hashem Pesaran
University of Southern California, and Trinity College, Cambridge

Qiankun Zhou†
State University of New York at Binghamton

June 11, 2016

Abstract

This paper proposes the Fixed Effects Filtered (FEF) and Fixed Effects Filtered instrumental variable (FEF-IV) estimators for estimation and inference in the case of time-invariant effects in static panel data models when N is large and T is fixed. The FEF-IV allows for endogenous time-invariant regressors but assumes that there exists a sufficient number of instruments for such regressors. It is shown that the FEF and FEF-IV estimators are $\sqrt{N}$-consistent, and asymptotically normally distributed. The FEF estimator is compared with the Fixed Effects Vector Decomposition (FEVD) estimator proposed by Plumper and Troeger (2007) and conditions under which the two estimators are equivalent are established. It is also shown that the variance estimator proposed for FEVD estimator is inconsistent and its use could lead to misleading inference. Alternative variance estimators are proposed for both FEF and FEF-IV estimators which are shown to be consistent under fairly general conditions. The small sample properties of the FEF and FEF-IV estimators are investigated by Monte Carlo experiments, and it is shown that FEF has smaller bias and RMSE, unless an intercept is included in the second stage of the FEVD procedure which renders the FEF and FEVD estimators identical. The FEVD procedure, however, results in substantial size distortions since it uses incorrect standard errors. In the case where some of the time-invariant regressors are endogenous, the FEF-IV procedure is compared with a modified version of Hausman and Taylor (1981) (HT) estimator. It is shown that both estimators perform well and have similar small sample properties. But the application of standard HT procedure, that incorrectly assumes a sub-set of time-varying regressors are uncorrelated with the individual effects, will yield biased estimates and significant size distortions.

Keywords: Static panel data models, time-invariant effects, endogenous time-invariant regressors, Monte Carlo experiments, fixed effects filtered estimators.

JEL classification: C01, C23, C33

*We would like to thank Alex Chudik, Ron Smith and two anonymous referees for helpful comments and suggestions.

†Corresponding to Qiankun Zhou, Department of Economics, State University of New York at Binghamton, Binghamton, NY 13902, USA, qzhou@binghamton.edu

1 Introduction

Identification and estimation of the effects of time-invariant regressors, such as the effects of race or gender is often the focus of panel data analysis, yet estimation procedures such as fixed effects (FE), that yield consistent estimates of the coefficients of time-varying regressors under fairly general conditions, cannot be used for estimation of the time-invariant effects, since the FE transformation eliminates all time-invariant regressors. As a result estimation of time-invariant effects has posed a challenge in panel data econometrics - namely how to carry out inference on time-invariant effects without making strong assumptions on the correlation between unobserved individual effects and the time-varying regressors.

For the estimation of time-invariant effects, Plumper and Troeger (2007) As we shall see, whilst the FEVD approach can be modified to yield consistent estimates of the time-invariant e (PT) propose the so called Fixed Effects Vector Decomposition (FEVD) through a three-step procedure.[1]ffects, the variance estimator proposed by PT for their estimator is not consistent. PT do not provide any formal statistical proofs to support their stated claims about the consistency of their estimator and its variance estimator. See Greene (2011a).

In the case where one or more of the time-invariant regressors are endogenous, an early pioneering contribution by Hausman and Taylor (1981) (HT) proposed using instrumental variables in the context of a pooled random coefficient panel data model. The instruments are obtained by assuming that known sub-sets of time-varying and time-invariant regressors are exogenous. HT also assumed that individual-specific effects as well as the idiosyncratic errors of the panel data model under consideration are serially uncorrelated and homoskedastic. Some of these assumptions are relaxed in the subsequent literature, but the main idea that sub-sets of time-varying and time invariant regressors are exogenous is typically maintained. See also Amemiya and MaCurdy (1986), Breusch et al. (1989), Im et al. (1999) and Baltagi and Bresson (2012).

In this paper, we consider a general static panel data model, which allows for an arbitrary degree of correlation between the time-varying covariates and the individual effects, and propose the fixed-effects filtered (FEF) estimation for the coefficients of the time-invariant regressors when the cross-sectional observation, N, is large and the time-series dimension, T, is small and fixed. Our proposed estimator has two simple steps. In the first step FE estimates are computed for the coefficients of the time-varying variables, and these estimates are used to filter out the time-varying effects. The residuals from the first stage panel regression are then averaged over time and used as a dependent variable in a cross-section OLS regression that includes an intercept and the vector of time-invariant regressors. Under the identifying assumption that the time-invariant regressors are uncorrelated with the individual effects and a number of other regularity conditions, it is shown that the FEF estimator is unbiased and consistent for a finite T and as $N \to \infty$. We derive the asymptotic distribution of the FEF estimator and propose a non-parametric estimator of its covariance matrix, not known in the literature, which we show to be consistent in the presence

[1] Plumper and Troeger (2007)'s FEVD approach is very popular in political science, and there even is a STATA procedure for the implementation of the FEVD estimator.

of heteroskedasticity of the individual effects and performs well in the presence of residual serial correlation.

Finally, we consider the case when one or more of the time-invariant variables are endogenous, and develop the FEF-IV estimator assuming there exist valid instruments. It is shown that the FEF-IV estimator is consistent and asymptotically normally distributed. A feasible variance estimator is also proposed for this FEF-IV estimator, which works well under heteroskedasticity and residual serial correlation.

The main advantage of the proposed FEF-IV over the HT estimator lies in the fact that it does not require a sub-set of time-varying regressors to be exogenous. This is in contrast to the HT procedure that uses a sub-set of time varying regressors as instruments for the time invariant regressors. This is a restrictive requirement since in panel data sets with fixed effects it is important to allow for all time-varying regressors to be correlated with the individual effects, and thus there may not exist any time-varying regressors to be used as instruments for the endogenous time-invariant regressors as required by the HT procedure. In such cases the HT procedure must be modified so that none of the time-varying regressors are used as instruments, the modification of HT estimation using exogenous time-invariant variables as instruments is also discussed in this paper, and comparison between the FEF-IV estimation and the modified HT estimation is also made. The second advantage of the FEF-IV estimator of time-invariant effects is its robustness to residual serial correlation and error heteroskedasticity.

We also contribute to the controversy over the FEVD estimator proposed by PT, discussed by Greene (2011a) and Breusch et al. (2011b), and followed up with responses and rejoinders by Plumper and Troeger (2011), Greene (2011b), Breusch et al. (2011a), and Beck (2011). The FEVD estimator of PT is based on a three step procedure, we show that when an intercept is included in the second step of their procedure, then the FEVD estimator is identical to the FEF estimator. But if an intercept is not included in the second stage, the FEVD estimator is in general biased and inconsistent. The extent of the bias of the FEVD estimator depends on the magnitude of intercept and the mean of time-invariant variables. More importantly, even if an intercept is included in the second step of the FEVD procedure, as we show, the variance of the FEVD estimator in the third step of PT's estimation procedure is biased and its use can lead to misleading inference. This is confirmed by the Monte Carlo simulations.

The small sample properties of the FEF and FEF-IV estimators for static panel data model are investigated, using two sets of comprehensive Monte Carlo experiments including error variance heteroskedasticity and residual serial correlation. In one set we generate the time-invariant regressors as exogenous, whilst in the second set we allow one of the time-invariant regressors to be correlated with the fixed effects. In both sets of experiments we allow the time-varying regressors to be correlated with the fixed effects. We compare FEF and FEVD estimators using the first set of experiments only, since these procedures are not appropriate in the case of the second set of experiments where one of the time-invariant regressors is endogenous. We find that our proposed estimator has smaller bias and RMSE, unless an intercept is included in the second stage of the

FEVD procedure which renders the FEF and FEVD estimators identical. However, as predicted by our theoretical derivations, the FEVD procedure results in substantial size distortions since it uses incorrect standard errors. In contrast, the use of the standard errors derived in this paper yields the correct size and satisfactory power in the case of all experiments, illustrating the robustness of our variance formula to heteroskedasticity and residual serial correlation. We also compare the FEF-IV estimator with the HT estimator using experiments where the time-invariant regressors are correlated with the fixed effects. The FEF-IV procedure performs well and has the correct size when appropriate instruments are used for the endogenous time-invariant regressors. Furthermore, it is robust to error variance heteroskedasticity and residual serial correlation. We also compare the FEF-IV estimator to a modified version of HT estimators and show that both estimators perform well and have similar small sample properties. But the application of standard HT procedure, that incorrectly assumes a sub-set of time-varying regressors are uncorrelated with the individual effects, will yield biased estimates and misleading inference.

The rest of the paper is organized as follows: Section 2 sets out the panel data model with time-invariant effects. Section 3 develops the FEF estimator, derives its asymptotic distribution, gives robust variance matrix estimator for the proposed FEF estimator, and provides a comparison of the FEF and FEVD estimators. Section 4 considers HT estimator as well as the modified HT estimation and extends the FEF estimator to the case when there are endogenous time-invariant regressors. Section 5 presents small sample properties of the FEF and FEF-IV estimators which are compared to FEVD, and modified HT estimators, respectively. The paper ends with some concluding remarks in Section 6. Some of the detailed mathematical proofs are provided in the Appendix. Additional Monte Carlo results, a derivation of the modified HT estimator and its comparison with the FEF-IV estimator are also provided in a Supplement which is available upon request.

Notations: For any real-valued $N \times N$ matrix $\mathbf{A}$, we will use $\|\mathbf{A}\|$ to denote the Frobenius norm of matrix $\mathbf{A}$ defined as $\|\mathbf{A}\| = [tr(\mathbf{A}\mathbf{A}')]^{1/2}$. Throughout, K denotes a generic non-zero positive constant that does not depend on N. The symbols $\rightarrow_p$ and $\rightarrow_d$ are used to denote convergence in probability and in distribution, respectively.

2 Panel data models with time-invariant effects

Consider the following panel data model that contains time-varying as well as time-invariant regressors

$$y_{it} = \alpha_i + \mathbf{z}_i'\boldsymbol{\gamma} + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \quad i = 1, 2, \ldots, N; t = 1, 2, \ldots, T, \tag{1}$$

where

$$\alpha_i = \alpha + \eta_i, \tag{2}$$

$\mathbf{x}_{it}$ is a $k \times 1$ vector of time-varying variables, and $\mathbf{z}_i$ is an $m \times 1$ vector of observed individual-specific variables that only vary over the cross section units, i. In addition to $\mathbf{z}_i$, the outcomes,

y_{it}, are also governed by unobserved individual specific effects, α_i. The focus of the analysis is on estimation and inference involving the elements of $\boldsymbol{\gamma}$. It is clear that without further restrictions on α_i, $\boldsymbol{\gamma}$ cannot be identified even if $\boldsymbol{\beta}$ was known to the researcher. For example consider the simple case where $\boldsymbol{\beta} = \mathbf{0}$, and assume that T is small. Then averaging across t we obtain

$$\bar{y}_i = \alpha + \mathbf{z}_i'\boldsymbol{\gamma} + v_i,$$

where $\bar{y}_i = T^{-1}\Sigma_{t=1}^{T} y_{it}$, $v_i = \eta_i + \bar{\varepsilon}_i$, and $\bar{\varepsilon}_i = T^{-1}\Sigma_{t=1}^{T}\varepsilon_{it}$. It is clear that without specifying how v_i and $\mathbf{z}_i$ are related it will not be possible to identify the effects of $\mathbf{z}_i$. To deal with this problem, it is often assumed that there exists instruments that are uncorrelated with v_i but at the same time are sufficiently correlated with $\mathbf{z}_i$. Even if such instruments exist a number of further complications arises if $\boldsymbol{\beta} \neq \mathbf{0}$. In such a case the IV approach must be extended also to deal with the possible dependencies between η_i and $\mathbf{x}_{it}$. In what follows we allow for η_i and $\mathbf{x}_{it}$ to have any degree of dependence, but initially assume that $\mathbf{z}_i$ and v_i are uncorrelated for identification of $\boldsymbol{\gamma}$, and assume that $\mathbf{x}_{it}$ and ε_{is} are uncorrelated for all i,t and s, to identify $\boldsymbol{\beta}$. This approach can be modified in cases where one or more instruments are available for $\mathbf{z}_i$ and/or $\mathbf{x}_{it}$.

3 Fixed effects filtered (FEF) estimator of time-invariant effects

3.1 FEF estimator

Under the assumption that $\mathbf{x}_{it}$ and $\boldsymbol{\varepsilon}_{is}$ are uncorrelated for all i, t and s, as it is well known $\boldsymbol{\beta}$ can be estimated consistently under fairly general assumptions on temporal dependence and cross-sectional heteroskedasticity of ε_{it}, and the distribution of the fixed effects, α_i. Denoting the FE estimator of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$, $\boldsymbol{\gamma}$ can then be estimated by the regression of $\bar{y}_i - \hat{\boldsymbol{\beta}}'\bar{\mathbf{x}}_i$ on an intercept and $\mathbf{z}_i$. We denote this estimator by $\hat{\boldsymbol{\gamma}}_{FEF}$ and refer to it as the fixed effects filtered (FEF) estimator of $\boldsymbol{\gamma}$. Formally, the FEF estimator can be computed using the following two-step procedure:

Step 1: Using model (1), compute the fixed-effects estimator of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}$, and the associated residuals $\hat{u}_{it}$ defined by

$$\hat{u}_{it} = y_{it} - \hat{\boldsymbol{\beta}}'\mathbf{x}_{it}. \tag{3}$$

Step 2: Compute the time averages of these residuals, $\bar{\hat{u}}_i = T^{-1}\sum_{t=1}^{T}\hat{u}_{it}$, and regress $\bar{\hat{u}}_i$ on $\mathbf{z}_i$ with an intercept to obtain $\hat{\boldsymbol{\gamma}}_{FEF}$, namely

$$\hat{\boldsymbol{\gamma}}_{FEF} = \left[\sum_{i=1}^{N}\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)'\right]^{-1}\sum_{i=1}^{N}\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)\left(\bar{\hat{u}}_i - \bar{\hat{u}}\right), \tag{4}$$

and

$$\hat{\alpha}_{FEF} = \bar{\hat{u}} - \hat{\boldsymbol{\gamma}}_{FEF}'\bar{\mathbf{z}}, \tag{5}$$

where $\bar{\hat{u}} = N^{-1}\sum_{i=1}^{N}\bar{\hat{u}}_i$.

The use of the FE residuals, $\hat{u}_{it}$, for consistent estimation of $\boldsymbol{\gamma}$ is not new and has been used in the literature extensively starting with the pioneering contribution of Hausman and Taylor (1981). The FEVD procedure proposed by Plumper and Troeger (2007) also makes use of the FE residuals. (see Section 3.4). The main contribution of this paper lies in development of the asymptotic distribution of $\hat{\boldsymbol{\gamma}}_{FEF}$ (and its IV version, $\hat{\boldsymbol{\gamma}}_{FEF-IV}$ introduced in Section 4.2) under fairly general conditions on the error processes ε_{it}, and η_i, and alternative assumptions concerning the correlation of $\mathbf{z}_i$ and $\eta_i + \bar{\varepsilon}_i$. We also derive conditions under which the covariance matrix of $\hat{\boldsymbol{\gamma}}_{FEF}$ (and $\hat{\boldsymbol{\gamma}}_{FEF-IV}$) can be consistently estimated.

3.2 Asymptotic Properties of the FEF Estimator of $\boldsymbol{\gamma}$

We examine the asymptotic properties of the FEF estimator of $\boldsymbol{\gamma}$, $\hat{\boldsymbol{\gamma}}_{FEF}$, defined by (4), under the following assumptions:

Assumption P1: $E(\varepsilon_{it} \left| \mathbf{x}_{is} \right.) = 0$, for all i, t and s, and $E\left(\varepsilon_{it}^4\right) < K < \infty$, for all i and t.

Assumption P2: $E\left(\varepsilon_{it}\varepsilon_{js} \left| \mathbf{X} \right.\right) = 0$, for all $i \neq j$, and all t and s, where $\mathbf{X} = (\mathbf{x}_{it}; i = 1, 2, ..., N;$ $t = 1, 2, ..., T)$.

Assumption P3: The errors, ε_{it}, are heteroskedastic and temporally dependent, namely

$$E\left(\varepsilon_{it}\varepsilon_{is} \left| \mathbf{X} \right.\right) = \gamma_i(t, s), \text{ for all } t \text{ and } s,$$

where $0 < \gamma_i(t, t) = \sigma_i^2$, and $|\gamma_i(t, s)| < K$, for all i, t and s.

Assumption P4 : The regressors, $\mathbf{x}_{it}$ satisfy the moment conditions $E\left\|\mathbf{x}_{it} - \bar{\mathbf{x}}\right\|^4 < K < \infty$, and $E\left\|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}\right\|^4 < K < \infty$, for all i and t, where $\bar{\mathbf{x}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{it}$, and $\bar{\mathbf{x}} = N^{-1}\sum_{i=1}^{N}\bar{\mathbf{x}}_i$.

Assumption P5: The $k \times k$ matrices $\mathbf{Q}_{p,NT}$ and $\mathbf{Q}_{FE,NT}$ defined by

$$\mathbf{Q}_{p,NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}\right)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}\right)', \tag{6}$$

$$\mathbf{Q}_{FE,NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)', \tag{7}$$

converge (in probability) to $\mathbf{Q}_{p,T}$ and $\mathbf{Q}_{FE,T}$ for a fixed T and as N tends to infinity, $\lambda_{\min}\left(\mathbf{Q}_{FE,NT}\right) >$ $1/K$ and $\lambda_{\min}\left(\mathbf{Q}_{p,NT}\right) > 1/K$, for all N and T where K is a finite, non-zero constant.

Assumption P6: The $m \times m$ matrix, $\mathbf{Q}_{zz,N}$, and the $m \times k$ matrix $\mathbf{Q}_{z\bar{x},N}$ defined by

$$\mathbf{Q}_{zz,N} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)', \tag{8}$$

$$\mathbf{Q}_{z\bar{x},N} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)\left(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}\right)', \tag{9}$$

converge (in probability) to the non-stochastic limits $\mathbf{Q}_{zz}$ and $\mathbf{Q}_{z\bar{x}}$, and $\lambda_{\min}\left(\mathbf{Q}_{zz,N}\right) > 1/K$, for all $N > m$.

Assumption P7: The time-invariant regressors, $\mathbf{z}_i$, are independently distributed of $v_j = \eta_j + \bar{\varepsilon}_j$, for all i and j, and η_i and $\bar{\varepsilon}_i$ are independently distributed. Also, $\mathbf{z}_i$ satisfy the moment conditions $E\left\|(\mathbf{z}_i - \bar{\mathbf{z}})\right\|^4 < K$, for all i.

Remark 1 *Note that since*

$$\left\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right\| = \left\|\mathbf{x}_{it} - \bar{\mathbf{x}} - (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})\right\| \leq \left\|\mathbf{x}_{it} - \bar{\mathbf{x}}\right\| + \left\|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}\right\|,$$

then any order moment conditions on $\|\mathbf{x}_{it} - \bar{\mathbf{x}}\|$ and $\|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}\|$ imply the same order moment conditions on $\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\|$. The boundedness of $\|\mathbf{x}_{it} - \bar{\mathbf{x}}\|$ and $\|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}\|$ are also sufficient for the boundedness of $\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\|$.

Remark 2 *Assumptions P5 and P6 ensure that there exists a finite N_0 such that for all $N > N_0$, $\mathbf{Q}_{zz,N}$ and $\mathbf{Q}_{FE,NT}$ are positive definite and converge in probability to the fixed matrices $\mathbf{Q}_{zz}$ and $\mathbf{Q}_{FE}$, respectively. But using the results in lemma A.1 in the Appendix, one can then relax the conditions $\lambda_{\min}(\mathbf{Q}_{zz,N}) > 1/K$ and $\lambda_{\min}(\mathbf{Q}_{FE,NT}) > 1/K$ by requiring $\lambda_{\min}(\mathbf{Q}_{zz}) > 2/K$ and $\lambda_{\min}(\mathbf{Q}_{FE}) > 2/K$. Under our assumptions the latter conditions ensure that the former conditions hold with probability approaching one.*

Remark 3 *Although, our focus is on fixed T and N large panels, we shall also discuss conditions under which our analysis will be valid when both T and N are large.*

The main result for the FEF estimator (4) is summarized in the following theorem.

Theorem 1 *Consider the FEF estimator $\hat{\boldsymbol{\gamma}}_{FEF}$ of $\boldsymbol{\gamma}$ in the panel data model (1) defined by (4), and suppose that Assumptions P1-P7 hold. Then $\hat{\boldsymbol{\gamma}}_{FEF}$ is an unbiased and a consistent estimator of $\boldsymbol{\gamma}$, and*

$$\sqrt{N}\left(\hat{\boldsymbol{\gamma}}_{FEF} - \boldsymbol{\gamma}\right) \rightarrow_d N\left(\mathbf{0}, \boldsymbol{\Omega}_{\hat{\gamma}_{FEF}}\right), \tag{10}$$

where

$$\boldsymbol{\Omega}_{\hat{\gamma}_{FEF}} = \mathbf{Q}_{zz}^{-1}\left(\sigma_\eta^2 \mathbf{Q}_{zz} + \boldsymbol{\Omega}_{\bar{\xi}}\right)\mathbf{Q}_{zz}^{-1}. \tag{11}$$

$\mathbf{Q}_{zz}$ is defined in Assumption P6, $\boldsymbol{\Omega}_{\bar{\xi}} = \lim_{N\to\infty} \boldsymbol{\Omega}_{\bar{\xi},N}$, with $\boldsymbol{\Omega}_{\bar{\xi},N}$ is defined by (A.24), and σ_η^2 is the variance of the fixed effects defined by (2).

Proof is provided in Section A.1 in the Appendix.

3.3 Consistent estimation of $Var\left(\hat{\boldsymbol{\gamma}}_{FEF}\right)$

In order to estimate $\boldsymbol{\Omega}_{\hat{\gamma}_{FEF}}$, it is helpful to begin with the following proposition regarding $\Omega_{\bar{\xi}}$, defined by (A.24), which enters the expression for $\boldsymbol{\Omega}_{\hat{\gamma}_{FEF}}$.

Proposition 1 *Let*

$$\mathbf{V}_{zz,N} = \frac{1}{N}\sum_{i=1}^{N} \omega_{iT}^2 \left(\mathbf{z}_i - \bar{\mathbf{z}}\right)\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)', \tag{12}$$

where $\omega_{iT}^2 = \sigma_\eta^2 + \kappa_{iT}^2$,

$$\kappa_{iT}^2 = \frac{\sigma_i^2}{T} + \frac{1}{T^2}\sum_{s\neq t}\gamma_i\left(s,t\right), \tag{13}$$

$\gamma_i\left(t,s\right) = E\left(\varepsilon_{it}\varepsilon_{is}\right)$, *and* $\sigma_i^2 = \gamma_i\left(t,t\right)$. *Then* $\sigma_\eta^2\mathbf{Q}_{zz} + \Omega_{\bar{\xi},N}$, *with* $\Omega_{\bar{\xi},N}$ *defined by (A.24), can be written as*

$$\sigma_\eta^2\mathbf{Q}_{zz} + \Omega_{\bar{\xi},N} = \mathbf{V}_{zz,N} + \mathbf{Q}_{z\bar{x},N}Var\left(\sqrt{N}\hat{\boldsymbol{\beta}}\right)\mathbf{Q}'_{z\bar{x},N} - \left(\boldsymbol{\Delta}_{\bar{\xi},N} + \boldsymbol{\Delta}'_{\bar{\xi},N}\right), \tag{14}$$

where $\mathbf{Q}_{z\bar{x},N}$ *is defined in (9), and*

$$\boldsymbol{\Delta}_{\bar{\xi},N} = \mathbf{Q}_{z\bar{x},N}\mathbf{Q}_{FE,NT}^{-1}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[T^{-2}\sum_{t,s=1}^{T}\gamma_i(t,s)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)'\right]\right\}. \tag{15}$$

Proof. A proof is provided in Section A.3 in the Appendix. ■

Proposition 2 *Under Assumptions P1-P7, and if*

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t,s=1}^{T}\gamma_i(t,s)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)' = o_p\left(1\right), \tag{16}$$

then the variance of the FEF estimator (4), can be consistently estimated for a fixed T *and as* $N \to \infty$, *by*

$$\hat{\boldsymbol{\Omega}}_{\hat{\gamma}_{FEF}} = N\ \widehat{Var}\left(\hat{\boldsymbol{\gamma}}_{FEF}\right) = \mathbf{Q}_{zz,N}^{-1}\left[\hat{\mathbf{V}}_{zz,N} + \mathbf{Q}_{z\bar{x},N}\left(N\widehat{Var}(\hat{\boldsymbol{\beta}})\right)\mathbf{Q}'_{z\bar{x},N}\right]\mathbf{Q}_{zz,N}^{-1}, \tag{17}$$

where $\mathbf{Q}_{zz,N}$ *and* $\mathbf{Q}_{z\bar{x},N}$ *are defined by (8) and (9), respectively,*

$$\widehat{Var}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{N}\mathbf{x}'_{i\cdot}\mathbf{x}_{i\cdot}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{x}'_{i\cdot}\mathbf{e}_i\mathbf{e}'_i\mathbf{x}_{i\cdot}\right)\left(\sum_{i=1}^{N}\mathbf{x}'_{i\cdot}\mathbf{x}_{i\cdot}\right)^{-1}, \tag{18}$$

where $\mathbf{x}'_{i\cdot} = \left(\mathbf{x}_{i1} - \bar{\mathbf{x}}_i, \mathbf{x}_{i2} - \bar{\mathbf{x}}_i, \ldots, \mathbf{x}_{iT} - \bar{\mathbf{x}}_i\right)$ *denotes the demeaned vector of* $\mathbf{x}_{it}$, *and the t-th element of* $\mathbf{e}_i$ *is given by* $e_{it} = y_{it} - \bar{y}_i - \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)'\hat{\boldsymbol{\beta}}$, *and*

$$\hat{\mathbf{V}}_{zz,N} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\varsigma}_i - \bar{\hat{\varsigma}}\right)^2(\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})', \tag{19}$$

where

$$\hat{\varsigma}_i - \bar{\hat{\varsigma}} = \bar{y}_i - \bar{y} - \left(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}\right)'\hat{\boldsymbol{\beta}} - (\mathbf{z}_i - \bar{\mathbf{z}})'\hat{\boldsymbol{\gamma}}_{FEF}. \tag{20}$$

Proof. A proof is provided in Section A.4 of the Appendix. ■

Condition (16) is not as restrictive as it may appear at first, and holds under a number of still fairly general assumptions regarding the error processes, ε_{it}. To see this, first not that

$$\frac{1}{NT^2}\sum_{t,s=1}^{T}\sum_{i=1}^{N}\gamma_i(t,s)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)' = T^{-1}\left[N^{-1}\sum_{i=1}^{N}\left(\frac{\mathbf{X}_i'\mathbf{M}_T\boldsymbol{\Gamma}_i\boldsymbol{\tau}_T\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'}{T}\right)\right],$$

where $\mathbf{X}_i = (\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})'$, $\boldsymbol{\Gamma}_i = (\gamma_i(t,s))$, and $\boldsymbol{\tau}_T$ is a $T\times 1$ vector of ones. Also, $T^{-1}\boldsymbol{\Gamma}_i\boldsymbol{\tau}_T = (\bar{\gamma}_{i1},\bar{\gamma}_{i2},...,\bar{\gamma}_{iT})'$, where $\bar{\gamma}_{it} = T^{-1}\sum_{s=1}^{T}\gamma_i(t,s)$. Then condition (16) is met exactly if $\bar{\gamma}_{it} = c_i$ for all t. Since in such a case $T^{-1}\boldsymbol{\Gamma}_i\boldsymbol{\tau}_T = c_i\boldsymbol{\tau}_T$, and $T^{-1}\mathbf{X}_i'\mathbf{M}_T\boldsymbol{\Gamma}_i\boldsymbol{\tau}_T\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)' = c_i\mathbf{X}_i'\mathbf{M}_T\boldsymbol{\tau}_T\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)' = 0$. Condition $\bar{\gamma}_{it} = c_i$ is clearly met if $\gamma_i(t,s) = 0$ for all $t\neq s$, and $\gamma_i(t,t) = E(\varepsilon_{it}^2) = \sigma_i^2$. After extensive simulations including cases where there are significant variations over time in $\bar{\gamma}_{it}$, we find that the effect of $\boldsymbol{\Delta}_{\bar{\xi},N}$ is negligible and the use of (17) for inference seems to be justified more generally.[2] Furthermore, the quality of approximating the variance of $\hat{\boldsymbol{\gamma}}_{FEF}$ by (17) tends to improve with T so long as $T^{-1}\sum_{t,s=1}^{T}|\gamma_i(t,s)| < K$.

3.4 Comparison of FEF and FEVD estimators

In this section, we will compare the FEF estimator with the FEVD proposed by Plumper and Troeger (2007). The FEVD procedure is based on the following three steps:

Step 1: The fixed effects approach is applied to (1), to compute the FE residuals, $\hat{u}_{it}$, defined by (3).

Step 2: In the second step, PT regress $\bar{\hat{u}}_i$ on $\mathbf{z}_i$ where $\bar{\hat{u}}_i = T^{-1}\sum_{t=1}^{T}\hat{u}_{it} = \bar{y}_i - \bar{\mathbf{x}}_i'\hat{\boldsymbol{\beta}}$, $\bar{y}_i = T^{-1}\sum_{t=1}^{T}y_{it}$ and $\bar{\mathbf{x}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{it}$. To obtain equivalence between the FEVD and FEF estimators of $\boldsymbol{\gamma}$, we modify this regression by also including an intercept in the regression and hence define the residuals from the second stage by

$$\hat{h}_i = \bar{\hat{u}}_i - \hat{a} - \mathbf{z}_i'\hat{\boldsymbol{\gamma}}, \tag{21}$$

where $\hat{a} = \bar{\bar{\hat{u}}} - \bar{\mathbf{z}}'\hat{\gamma}$, $\bar{\bar{\hat{u}}} = N^{-1}\sum_{i=1}^{N}\bar{\hat{u}}_i$, and

$$\hat{\boldsymbol{\gamma}} = \left[\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\right]^{-1}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\bar{\hat{u}}_i-\bar{\bar{\hat{u}}}\right). \tag{22}$$

which is exactly the same as our FEF estimator. Using the above results we now have

$$\begin{aligned}\hat{h}_i - \bar{\hat{h}} &= \left(\bar{\hat{u}}_i - \bar{\bar{\hat{u}}}\right) - \left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\hat{\boldsymbol{\gamma}} \\ &= \left(\bar{y}_i - \bar{y}\right) - \left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)'\hat{\boldsymbol{\beta}} - \left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\hat{\boldsymbol{\gamma}},\end{aligned} \tag{23}$$

[2] The results of these simulations are available upon request.

where $\bar{\hat{h}} = N^{-1}\sum_{i=1}^{N}\hat{h}_i$. Also, from the normal equations of this step, note that

$$\bar{\hat{h}} = 0,\ N^{-1}\sum_{i=1}^{N}\left(\hat{h}_i - \bar{\hat{h}}\right)(\mathbf{z}_i - \bar{\mathbf{z}})' = \mathbf{0}. \tag{24}$$

Step 3: The third step uses $\hat{h}_i$ computed in the earlier stage, as defined by (21), and estimates the following panel regression by pooled OLS

$$y_{it} = a + \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\gamma} + \delta\hat{h}_i + \tilde{\varepsilon}_{it}. \tag{25}$$

These estimators are the modified FEVD estimators which we shall denote by $\tilde{\boldsymbol{\gamma}}$, $\tilde{\boldsymbol{\delta}}$ and $\tilde{\boldsymbol{\beta}}$, and as before we denote the FE estimator of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$, and the estimator of $\boldsymbol{\gamma}$ obtained in the second step of FEVD approach by $\hat{\boldsymbol{\gamma}}$ (which is identical to the FEF estimator if an intercept is included in the second step). The original FEVD estimators proposed by PT are based on the same pooled OLS regression, but do not include an intercept in the second stage regression that computes the $\hat{h}_i$.[3] As we shall see this makes a great deal of difference to the resultant estimators.

To investigate the relationship between FEF and FEVD estimators we first introduce the following notations

$$\begin{aligned}
\mathbf{q}_{p,NT} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}})(y_{it} - \bar{y}),\ \mathbf{Q}_{z\bar{y},N} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \bar{\mathbf{z}})(\bar{y}_i - \bar{y}), && (26)\\
\mathbf{Q}_{h\bar{x},N} &= \frac{1}{N}\sum_{i=1}^{N}\left(\hat{h}_i - \bar{\hat{h}}\right)(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})',\ \mathbf{Q}_{hz,N} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{h}_i - \bar{\hat{h}}\right)(\mathbf{z}_i - \bar{\mathbf{z}})', && (27)\\
\mathbf{Q}_{h\bar{y},N} &= \frac{1}{N}\sum_{i=1}^{N}\left(\hat{h}_i - \bar{\hat{h}}\right)(\bar{y}_i - \bar{y}),\ \mathbf{Q}_{hh,N} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{h}_i - \bar{\hat{h}}\right)^2, && (28)
\end{aligned}$$

where $\bar{\hat{h}}$, $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}_i$, $\bar{\mathbf{z}}$, $\bar{y}$, and $\bar{y}_i$ are defined as before. Using these additional notations, the normal equations of the pooled OLS regressions for the panel data model defined by (25) are given by

$$\begin{aligned}
\mathbf{q}_{p,NT} &= \mathbf{Q}_{p,NT}\tilde{\boldsymbol{\beta}} + \mathbf{Q}_{z\bar{x},N}'\tilde{\boldsymbol{\gamma}} + \mathbf{Q}_{h\bar{x},N}'\tilde{\delta},\\
\mathbf{Q}_{z\bar{y},N} &= \mathbf{Q}_{z\bar{x},N}\tilde{\boldsymbol{\beta}} + \mathbf{Q}_{zz,N}\tilde{\boldsymbol{\gamma}} + \mathbf{Q}_{hz,N}'\tilde{\delta},\\
\mathbf{Q}_{h\bar{y},N} &= \mathbf{Q}_{h\bar{x},N}\tilde{\boldsymbol{\beta}} + \mathbf{Q}_{hz,N}\tilde{\boldsymbol{\gamma}} + \mathbf{Q}_{hh,N}\tilde{\delta}.
\end{aligned}$$

Also, when an intercept is included in the second step of FEVD we have $\mathbf{Q}_{hz,N} = 0 = \mathbf{Q}_{hz,N}'$, (see

[3] For example, equations (4) and (5) on p128 of Plumper and Troeger (2007).

(24) and (27)), and the normal equations reduce to

$$\mathbf{q}_{p,NT} = \mathbf{Q}_{p,NT}\tilde{\boldsymbol{\beta}} + \mathbf{Q}'_{z\bar{x},N}\tilde{\boldsymbol{\gamma}} + \mathbf{Q}'_{h\bar{x},N}\tilde{\delta}, \tag{29}$$

$$\mathbf{Q}_{z\bar{y},N} = \mathbf{Q}_{z\bar{x},N}\tilde{\boldsymbol{\beta}} + \mathbf{Q}_{zz,N}\tilde{\boldsymbol{\gamma}}, \tag{30}$$

$$\mathbf{Q}_{h\bar{y},N} = \mathbf{Q}_{h\bar{x},N}\tilde{\boldsymbol{\beta}} + \mathbf{Q}_{hh,N}\tilde{\delta}. \tag{31}$$

The FEVD estimator of $\boldsymbol{\gamma}$, namely $\tilde{\boldsymbol{\gamma}}$, can now be obtained using the above system of the equations. The results are summarized in the following proposition.

Proposition 3 *Consider the panel data model (25), and suppose that $\mathbf{Q}_{p,NT}$ and $\mathbf{Q}_{zz,N}$ are non-singular, and $\mathbf{Q}_{hh,N} > 0$. Let*

$$\mathbf{Q}_{NT} = \mathbf{Q}_{p,NT} - \mathbf{Q}'_{z\bar{x},N}\mathbf{Q}^{-1}_{zz,N}\mathbf{Q}_{z\bar{x},N} - \mathbf{Q}'_{h\bar{x},N}\mathbf{Q}^{-1}_{hh,N}\mathbf{Q}_{h\bar{x},N}, \tag{32}$$

and suppose also that $\mathbf{Q}_{NT}$ is non-singular. Then the FEVD estimators proposed by PT and FEF estimators proposed in this paper are identical if an intercept is included in the second step regression of the FEVD procedure, namely $\tilde{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}$, and $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$. Furthermore $\tilde{\delta}$, the FEVD estimator of δ in the third step of the FEVD procedure, is identically equal to unity.

Proof. See Section A.5 in the Appendix for a proof. ■

Proposition 4 *Suppose that the three-step FEVD estimators (denoted as before by $\tilde{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\delta}}$) are computed without including an intercept in the regression in the second step. In this case we continue to have $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$ and $\tilde{\delta} = 1$, but for $\boldsymbol{\gamma}$ we obtain $\tilde{\boldsymbol{\gamma}} = \mathring{\boldsymbol{\gamma}}$, where $\mathring{\boldsymbol{\gamma}}$ is the OLS estimator of the coefficient of $\mathbf{z}_i$ in the OLS regression of $\bar{\hat{u}}_i$ on $\mathbf{z}_i$, without an intercept, and $\mathring{\boldsymbol{\gamma}}$ is biased and inconsistent unless $\alpha E\left[\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\right] = \mathbf{0}$.*

Proof. See Section A.6 in the Appendix for a proof. ■

It is also interesting to compare the covariance of the FEF given by (11), with the one that is obtained when the standard formula for the variance of the pooled OLS estimators is applied to the third step of the FEVD procedure as proposed by PT. Recall that the FEVD estimator of $\boldsymbol{\gamma}$ coincides with $\hat{\boldsymbol{\gamma}}$ if an intercept is included in the second step of the procedure, and pooled OLS applied to the third step will result in a valid inference only if the variance obtained using the FEVD procedure also coincides with $\boldsymbol{\Omega}_{\hat{\gamma}_{FEF}}$. To simplify the comparisons suppose that $\varepsilon_{it} \sim IID(0, \sigma^2)$ for all i and t, and note that in this simple case $\mathbf{V}_{zz}$ and $\boldsymbol{\Omega}_{\hat{\beta}}$ (given by (12) and (A.10)) reduce to

$$\mathbf{V}_{zz} = \left(\sigma_\eta^2 + \frac{\sigma^2}{T}\right)\mathbf{Q}_{zz},$$

$$\boldsymbol{\Omega}_{\hat{\beta}} = \frac{\sigma^2}{T}\left[\lim_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)'\right]^{-1} = \frac{\sigma^2}{T}\left(\mathbf{Q}_p - \mathbf{Q}_{\bar{x}\bar{x}}\right)^{-1},$$

and we have[4]

$$\begin{aligned}\mathbf{\Omega}_{\hat{\gamma}_{FEF}} &= \mathbf{Q}_{zz}^{-1}\left[\left(\sigma_\eta^2+\frac{\sigma^2}{T}\right)\mathbf{Q}_{zz}+\mathbf{Q}_{z\bar{x}}\Omega_{\hat{\beta}}\mathbf{Q}_{\bar{x}z}\right]\mathbf{Q}_{zz}^{-1} \\ &= \left(\sigma_\eta^2+\frac{\sigma^2}{T}\right)\mathbf{Q}_{zz}^{-1}+\frac{\sigma^2}{T}\mathbf{Q}_{zz}^{-1}\mathbf{Q}_{z\bar{x}}\left(\mathbf{Q}_p-\mathbf{Q}_{\bar{x}\bar{x}}\right)^{-1}\mathbf{Q}_{\bar{x}z}\mathbf{Q}_{zz}^{-1}.\end{aligned}$$

Under the same model specifications the covariance of the FEVD estimator (also scaled by $\sqrt{N}$) is given by

$$\mathbf{\Omega}_{\hat{\gamma}_{FEVD}}=\sigma_{\tilde{\varepsilon}}^2\left[\mathbf{Q}_{zz}^{-1}+\mathbf{Q}_{zz}^{-1}\mathbf{Q}_{z\bar{x}}\mathbf{Q}^{-1}\mathbf{Q}_{z\bar{x}}'\mathbf{Q}_{zz}^{-1}\right],$$

where as before $\mathbf{Q}=\mathbf{Q}_p-\mathbf{Q}_{z\bar{x}}'\mathbf{Q}_{zz}^{-1}\mathbf{Q}_{z\bar{x}}-\mathbf{Q}_{h\bar{x}}'\mathbf{Q}_{hh}^{-1}\mathbf{Q}_{h\bar{x}}$,

$$\begin{aligned}\sigma_{\tilde{\varepsilon}}^2 &= \lim_{N\to\infty}\left[N^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(y_{it}-\tilde{\alpha}-\mathbf{x}_{it}'\tilde{\boldsymbol{\beta}}-\mathbf{z}_i'\tilde{\boldsymbol{\gamma}}-\tilde{\delta}\hat{h}_i\right)^2\right] \\ &= \lim_{N\to\infty}\left[N^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\widehat{\tilde{\varepsilon}}_{it}^2\right],\end{aligned}$$

and

$$\widehat{\tilde{\varepsilon}}_{it}=y_{it}-\bar{y}-\left(\mathbf{x}_{it}-\bar{\mathbf{x}}\right)'\hat{\boldsymbol{\beta}}-\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\hat{\boldsymbol{\gamma}}-\hat{h}_i.$$

To derive $\sigma_{\tilde{\varepsilon}}^2$ note that $\hat{h}_i=\bar{y}_i-\bar{y}-(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\hat{\boldsymbol{\beta}}-(\mathbf{z}_i-\bar{\mathbf{z}})'\hat{\boldsymbol{\gamma}}$, and hence

$$\begin{aligned}\widehat{\tilde{\varepsilon}}_{it} &= y_{it}-\bar{y}-\left(\mathbf{x}_{it}-\bar{\mathbf{x}}\right)'\hat{\boldsymbol{\beta}}-\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\hat{\boldsymbol{\gamma}}-\hat{h}_i \\ &= y_{it}-\bar{y}_i-\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)'\hat{\boldsymbol{\beta}} \\ &= \varepsilon_{it}-\bar{\varepsilon}_i-\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right).\end{aligned}$$

Also, in the case where $\varepsilon_{it}\sim IID(0,\sigma^2)$, we have $Cov\left[\varepsilon_{it}-\bar{\varepsilon}_i,\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right]=\mathbf{0}$, and hence

$$E\left(\widehat{\tilde{\varepsilon}}_{it}^2\right)=\sigma^2\left(1-\frac{1}{T}+\frac{k}{N}\right).$$

A comparison of the expressions derived above for $\mathbf{\Omega}_{\hat{\gamma}_{FEF}}$ and $\mathbf{\Omega}_{\hat{\gamma}_{FEVD}}$ clearly shows that they differ irrespective of whether T is fixed or $T\to\infty$.

Remark 4 *Plumper and Troeger (2007) argue that the necessity of third step is to correct standard errors of* $\mathring{\boldsymbol{\gamma}}$ *(Plumper and Troeger (2007) , p129)***,** *however, as shown above and in the simulations below, the variance term calculated in the third step of FEVD does not fully correct the bias of the variance estimator in the second step.*[5]

[4] Note that since in the present case $\varepsilon_{it}'s$ are serially uncorrelated then $\mathbf{\Delta}_{\bar{\xi}}=0$.
[5] PT state "...only the third stage allows obtaining the correct SE's.", P.129.

4 Panel data models with endogenous time-invariant regressors

Plumper and Troeger (2007) only consider the case when the time-invariant regressors are exogenous, and their estimation procedure does not apply to the case where one or more time-invariant regressors are endogenous. In this case the two-step procedure proposed by Hausman and Taylor (1981) (HT) can be used, so long there are a sufficient number of time-varying regressors that are uncorrelated with the individual effects. In this section we begin by providing an overview of HT estimation procedure and propose an IV version of FEF estimator (denoted as FEF-IV) that allows for endogeneity of the time-invariant regressors, and does not require any of the time-varying regressors to be uncorrelated with the individual effects. We also propose a modified version of the HT estimator (denoted as HTM) that does not make use of time-varying regressors as instruments, and is comparable to our proposed FEF-IV estimator.

4.1 Hausman and Taylor estimator and its modification

Hausman and Taylor (1981) approach the problem of estimation of the time-invariant effects in the panel data model, (1), by assuming that $\mathbf{x}_{it}$ and $\mathbf{z}_i$ can be partitioned into two parts as $(\mathbf{x}_{1,it}, \mathbf{x}_{2,it})$ and $(\mathbf{z}_{1,i}, \mathbf{z}_{2,i})$, respectively, such that the following moment conditions hold

$$E\left(\mathbf{x}_{1,it}'\eta_i\right) = \mathbf{0},\ E\left(\mathbf{z}_{1,i}'\eta_i\right) = \mathbf{0}, \tag{33}$$

$$E\left(\mathbf{x}_{2,it}'\eta_i\right) \neq \mathbf{0},\ E\left(\mathbf{z}_{2,i}'\eta_i\right) \neq \mathbf{0}, \tag{34}$$

and $\dim\left(\mathbf{x}_{1,it}\right) \geq \dim\left(\mathbf{z}_{2,i}\right)$. To compute the HT estimator, the panel data model is first written as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \left(\mathbf{z}_i'\boldsymbol{\gamma} + \alpha + \eta_i\right)\boldsymbol{\tau}_T + \boldsymbol{\varepsilon}_i, \text{ for } i = 1, 2, ..., N, \tag{35}$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})'$, $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})'$. Then the following two-step procedure is used[6]:

Step 1 of HT: As in our approach, $\boldsymbol{\beta}$ is estimated by $\hat{\boldsymbol{\beta}}$, the FE estimator, and the deviations $\hat{d}_i = \bar{y}_i - \bar{\mathbf{x}}_i'\hat{\boldsymbol{\beta}}$, $i = 1, 2, \ldots, N$ are computed.

Step 2 of HT: In the second step, the deviations $\hat{\mathbf{d}} = (\hat{d}_1, \hat{d}_2, ..., \hat{d}_N)'$ are used to compute the 2SLS (or IV) estimator of $\boldsymbol{\gamma}$

$$\hat{\gamma}_{IV} = \left(\mathbf{Z}'\mathbf{P}_A\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{P}_A\hat{\mathbf{d}}, \tag{36}$$

where $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_N)' = (\mathbf{Z}_1, \mathbf{Z}_2)$, $\mathbf{P}_A = \mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'$ is the orthogonal projection matrix of $\mathbf{A} = \left(\boldsymbol{\tau}_N, \bar{\mathbf{X}}_1, \mathbf{Z}_1\right)$, where $\bar{\mathbf{X}} = (\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2)$, $\bar{\mathbf{X}} = (\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, ..., \bar{\mathbf{x}}_N)'$, and $\bar{\mathbf{x}}_i = (\bar{\mathbf{x}}_{i,1}, \bar{\mathbf{x}}_{i,2})$. Using these

[6] As noted in the Introduction, HT procedure is further developed and extended in the papers by Amemiya and MaCurdy (1986), Breusch et al. (1989), Im et al. (1999) and Baltagi and Bresson (2012).

initial estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the error variances σ_ε^2 and σ_η^2 are also estimated as

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)} \sum_{i=1}^{N} \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{FE}\right)' \mathbf{M}_T \left(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{FE}\right), \quad (37)$$

$$\hat{\sigma}_\eta^2 = \hat{s}^2 - \hat{\sigma}_\varepsilon^2, \quad (38)$$

$$\hat{s}^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \hat{\mu} - \mathbf{x}_{it}' \hat{\boldsymbol{\beta}}_{FE} - \mathbf{z}_i' \hat{\boldsymbol{\gamma}}_{IV}\right)^2. \quad (39)$$

Step 3 of HT: In the third step the N equations in (35) are stacked to obtain

$$\mathbf{y} = \mathbf{W}\boldsymbol{\theta} + (\boldsymbol{\eta} \otimes \boldsymbol{\tau}_T) + \boldsymbol{\varepsilon},$$

where $\mathbf{W} = [(\boldsymbol{\tau}_N \otimes \boldsymbol{\tau}_T), \mathbf{X}, (\mathbf{Z} \otimes \boldsymbol{\tau}_T)]$, $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}', \boldsymbol{\gamma}')'$, $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', ..., \mathbf{y}_N')'$, $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_N)'$, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', ..., \boldsymbol{\varepsilon}_N')'$. Under the assumptions that the errors are cross-sectionally independent, *serially uncorrelated* and *homoskedastic* we have

$$\boldsymbol{\Omega} = Var\left[(\boldsymbol{\eta} \otimes \boldsymbol{\tau}_T) + \boldsymbol{\varepsilon}\right] = \sigma_\eta^2 \left(\mathbf{I}_N \otimes \boldsymbol{\tau}_T \boldsymbol{\tau}_T'\right) + \sigma_\varepsilon^2 \left(\mathbf{I}_N \otimes \mathbf{I}_T\right),$$

which can be written as $\boldsymbol{\Omega} = \left(\sigma_\varepsilon^2 + T\sigma_\eta^2\right) \mathbf{P}_V + \sigma_\varepsilon^2 \mathbf{Q}_V$, where $\mathbf{P}_V = \mathbf{I}_N \otimes (\mathbf{I}_T - \mathbf{M}_T)$ and $\mathbf{Q}_V = \mathbf{I}_N \otimes \mathbf{M}_T$. It is now easily verified that

$$\boldsymbol{\Omega}^{-1/2} = \frac{1}{\sigma_\varepsilon} \left(\varphi \mathbf{P}_V + \mathbf{Q}_V\right) \quad (40)$$

where $\varphi = \sigma_\varepsilon / \sqrt{\sigma_\varepsilon^2 + T\sigma_\eta^2}$. Then the transformed model can be written as

$$\boldsymbol{\Omega}^{-1/2}\mathbf{y} = \boldsymbol{\Omega}^{-1/2}\mathbf{W}\boldsymbol{\theta} + \boldsymbol{\Omega}^{-1/2}\left[(\boldsymbol{\eta} \otimes \boldsymbol{\tau}_T) + \boldsymbol{\varepsilon}\right]. \quad (41)$$

To simplify the notations we assume that the first column of $\mathbf{Z}$ is $\boldsymbol{\tau}_N$, and then write the (infeasible) HT estimator as,

$$\hat{\boldsymbol{\theta}}_{HT} = \left(\mathbf{W}'\boldsymbol{\Omega}^{-1/2}\mathbf{P}_A\boldsymbol{\Omega}^{-1/2}\mathbf{W}\right)^{-1} \left(\mathbf{W}'\Omega^{-1/2}\mathbf{P}_A\Omega^{-1/2}\mathbf{y}\right), \quad (42)$$

where

$$\mathbf{P}_A = \mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}', \ \mathbf{A} = \left(\boldsymbol{\tau}_N \otimes \boldsymbol{\tau}_T, \mathbf{Q}_V\mathbf{X}, \mathbf{X}^{(1)}, \mathbf{Z}_1 \otimes \boldsymbol{\tau}_T\right),$$

$\mathbf{X}^{(1)} = (\mathbf{X}_{1,1}, \mathbf{X}_{1,2}, ..., \mathbf{X}_{1,N})'$, with $\mathbf{X}_{1,i}' = (\mathbf{x}_{1,i1}, \ldots, \mathbf{x}_{1,iT})$, and $\mathbf{x}_{1,it}$ contains the regressors that are uncorrelated with η_i.[7]

The variance covariance matrix of $\hat{\boldsymbol{\theta}}_{HT}$ in the general case where the fixed effects, η_i, are

[7] See Amemiya and MaCurdy (1986) and Breusch et al. (1989) for discussion on the choice of instruments for HT estimation.

heteroskedastic and possibly cross-sectionally correlated is given by [8]

$$Var\left(\hat{\boldsymbol{\theta}}_{HT}\right)=\mathbf{Q}^{-1}+\left(\frac{T}{\sigma_{\varepsilon}^{2}+T\sigma_{\eta}^{2}}\right)\mathbf{Q}^{-1}\left[\mathbf{W}'\boldsymbol{\Omega}^{-1/2}\mathbf{P}_{A}\left(\left(\mathbf{V}_{\eta}-\sigma_{\eta}^{2}\mathbf{I}_{N}\right)\otimes\frac{1}{T}\boldsymbol{\tau}_{T}\boldsymbol{\tau}_{T}'\right)\mathbf{P}_{A}\boldsymbol{\Omega}^{-1/2}\mathbf{W}\right]\mathbf{Q}^{-1}, \tag{43}$$

where $\mathbf{Q}=\mathbf{W}'\boldsymbol{\Omega}^{-1/2}\mathbf{P_A}\boldsymbol{\Omega}^{-1/2}\mathbf{W}$, and $\mathbf{V}_{\eta}$ represents the covariance matrix of $\boldsymbol{\eta}$. $Var\left(\hat{\boldsymbol{\theta}}_{HT}\right)$ reduces to $\mathbf{Q}^{-1}$ in the standard case where η_i's are assumed to be homoskedastic and cross-sectionally independent, namely when $\mathbf{V}_{\eta}=\sigma_{\eta}^{2}\mathbf{I}_{N}$. To our knowledge the above general expression for $Var\left(\hat{\boldsymbol{\theta}}_{HT}\right)$ is new.

The above HT estimator assumes that the errors η_i and ε_{it} are both homoskedastic, serially uncorrelated and cross-sectionally independent. However, simulations to be reported below suggest that the HT estimator works well even in cases where the errors are heteroskedastic and possibly serially correlated, assuming that the conditions (33) are satisfied. In view of this, and for comparability with the FEF-IV to be presented below, we also consider the following modified version of the HT estimator which is valid even if $E\left(\mathbf{x}_{1,it}'\eta_i\right)\neq\mathbf{0}$, but instead there exist instruments (possibly from outside the specified model) for the endogenous time-invariant regressors. The modified HT (HTM) estimator is computed assuming that there exists the $s\times 1$ vector of time-invariant instrumental variables, denoted by $\mathbf{r}_i$ such that $s\geq\dim\left(\mathbf{z}_{2,i}\right)$. In the first step the IV estimator of $\boldsymbol{\gamma}$ is computed by

$$\tilde{\boldsymbol{\gamma}}_{IV}=\left(\mathbf{Z}'\mathbf{P}_{H}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{P}_{H}\hat{\mathbf{d}}, \tag{44}$$

where $\hat{\mathbf{d}}$ and $\mathbf{Z}$ are as in the Step 1 of HT, $\mathbf{P}_{H}=\mathbf{H}\left(\mathbf{H}'\mathbf{H}\right)^{-1}\mathbf{H}'$ is the orthogonal projection matrix of instruments $\mathbf{H}=\left(\boldsymbol{\tau}_N,\mathbf{R}\right)$, where $\mathbf{R}=\left(\mathbf{r}_1,\ldots,\mathbf{r}_N\right)'$ is the N In the second step the (infeasible) HTM estimator is computed as

$$\hat{\boldsymbol{\theta}}_{HTM}=\left(\mathbf{W}'\boldsymbol{\Omega}^{-1/2}\mathbf{P}_{A^*}\boldsymbol{\Omega}^{-1/2}\mathbf{W}\right)^{-1}\left(\mathbf{W}'\Omega^{-1/2}\mathbf{P}_{A^*}\Omega^{-1/2}\mathbf{y}\right), \tag{45}$$

where $\boldsymbol{\Omega}$ is given by Step 2 of HT, $\mathbf{P}_{A^*}=\mathbf{A}^*\left(\mathbf{A}^{*\prime}\mathbf{A}^*\right)^{-1}\mathbf{A}^{*\prime}$ with $\mathbf{A}^*=\left[\left(\boldsymbol{\tau}_N\otimes\boldsymbol{\tau}_T\right),\mathbf{R}\otimes\boldsymbol{\tau}_T,\mathbf{Q}_V\mathbf{X}\right]$. The associated variance-covariance matrix is given by

$$Cov\left(\hat{\boldsymbol{\theta}}_{HTM}\right)=\left(\mathbf{W}'\boldsymbol{\Omega}^{-1/2}\mathbf{P}_{A^*}\boldsymbol{\Omega}^{-1/2}\mathbf{W}\right)^{-1}. \tag{46}$$

Feasible estimators of $\hat{\boldsymbol{\theta}}_{HTM}$ and $Cov\left(\hat{\boldsymbol{\theta}}_{HTM}\right)$ are obtained by replacing σ_ε and σ_η in $\boldsymbol{\Omega}^{-1/2}$ with their estimates given by (37) and (38), respectively, where $\hat{s}^2$ (defined by (39) is now computed using $\tilde{\boldsymbol{\gamma}}_{IV}$ given by (44) instead of $\hat{\boldsymbol{\gamma}}_{IV}$ given by (36).

4.2 FEF-IV estimation of time-invariant effects

Having introduced the modified HT estimator, we now provide a generalization of the FEF estimator that allows for possible endogeneity of the time-invariant regressors, assuming the existence of the

[8]See Section A.7 of the Appendix for a derivation.

same vector of time-invariant instruments. $\mathbf{r}_i$, as used in the modified HT set up above. We denote this estimator by FEF-IV and derive its asymptotic properties under the following assumptions that allow for the errors to be heteroskedastic and serially correlated:

Assumption P8: There exists the $s \times 1$ vector of instruments $\mathbf{r}_i$ for $\mathbf{z}_i$, $i = 1, 2, \ldots, N$, where $\mathbf{r}_i$ is distributed independently of η_j and $\bar{\varepsilon}_j$ for all i and j, and $s \geq m$, and satisfy the moment condition $E \left\| \mathbf{r}_i - \bar{\mathbf{r}} \right\|^4 < K < \infty$, if it has unbounded support.

Assumption P9: Let

$$\mathbf{Q}_{rz,N} = N^{-1} \sum_{i=1}^{N} (\mathbf{r}_i - \bar{\mathbf{r}})(\mathbf{z}_i - \bar{\mathbf{z}})', \ \mathbf{Q}_{r\bar{x},N} = N^{-1} \sum_{i=1}^{N} (\mathbf{r}_i - \bar{\mathbf{r}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})', \ \ \mathbf{Q}_{rr,N} = N^{-1} \sum_{i=1}^{N} (\mathbf{r}_i - \bar{\mathbf{r}})(\mathbf{r}_i - \bar{\mathbf{r}})', \tag{47}$$

where $\bar{\mathbf{r}} = N^{-1} \sum_{i=1}^{N} \mathbf{r}_i$, $\bar{\mathbf{z}} = N^{-1} \sum_{i=1}^{N} \mathbf{z}_i$,and $\bar{\mathbf{x}} = N^{-1} \sum_{i=1}^{N} \mathbf{x}_i$. The $s \times m$ matrix $\mathbf{Q}_{rz,N}$, and the $s \times s$ matrix $\mathbf{Q}_{rr,N}$ have full column ranks (namely $rank\left(\mathbf{Q}_{rz,N}\right) = m \leq s$ and $rank\left(\mathbf{Q}_{rr,N}\right) = s$), and have finite probability limits as $N \to \infty$ given by $\mathbf{Q}_{rz}$ and $\mathbf{Q}_{rr}$, respectively. Matrices $\mathbf{Q}_{r\bar{x},N}$ and $\mathbf{Q}_{zz,N}$ have finite probability limits given by $\mathbf{Q}_{r\bar{x}}$ and $\mathbf{Q}_{zz}$, respectively, and in cases where $\mathbf{x}_{it}$ and $\mathbf{z}_i$ are stochastic with unbounded supports, $1/K < \lambda_{\min}(\mathbf{Q}_{rr,N}) < \lambda_{\max}(\mathbf{Q}_{rr,N}) < K$, for some constant K, and for all N, and as $N \to \infty$, with probability approaching unity.

Under the above assumptions and maintaining the earlier Assumptions P1-P6, a consistent estimator of $\boldsymbol{\gamma}$ can be obtained as follows[9]

$$\hat{\boldsymbol{\gamma}}_{FEF-IV} = \left(\mathbf{Q}_{zr,N} \mathbf{Q}_{rr,N}^{-1} \mathbf{Q}_{zr,N}' \right)^{-1} \left(\mathbf{Q}_{zr,N} \mathbf{Q}_{rr,N}^{-1} \mathbf{Q}_{r\hat{u},N} \right), \tag{48}$$

where $\mathbf{Q}_{zr,N}$ and $\mathbf{Q}_{rr,N}$ are defined by (47),

$$\mathbf{Q}_{r\hat{u},N} = \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{r}_i - \bar{\mathbf{r}} \right) \left(\bar{\hat{u}}_i - \bar{\hat{u}} \right),$$

$\bar{\hat{u}} = \frac{1}{N} \sum_{i=1}^{N} \bar{\hat{u}}_i$, $\bar{\hat{u}}_i = \bar{y}_i - \left(\bar{\mathbf{x}}_i - \bar{\mathbf{x}} \right)' \hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\beta}}$ is the FE estimator of $\boldsymbol{\beta}$. It then follows that

$$\sqrt{N} \left(\hat{\boldsymbol{\gamma}}_{FEF-IV} - \boldsymbol{\gamma} \right) = \left(\mathbf{Q}_{zr,N} \mathbf{Q}_{rr,N}^{-1} \mathbf{Q}_{zr,N}' \right)^{-1} \left(\mathbf{Q}_{zr,N} \mathbf{Q}_{rr,N}^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\mathbf{r}_i - \bar{\mathbf{r}} \right) \zeta_i \right),$$

with $\zeta_i = \eta_i + \bar{\varepsilon}_i - \bar{\mathbf{x}}_i' \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right)$, as before (see (A.12)). Following a similar line of proof as in the case with exogenous $\mathbf{z}_i$, under Assumptions P1-P6 and P8-P9, it can be shown that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\mathbf{r}_i - \bar{\mathbf{r}} \right) \left(\zeta_i - \bar{\zeta} \right) \to_d N \left(\mathbf{0}, \sigma_\eta^2 \mathbf{Q}_{rr} + \boldsymbol{\Omega}_{\bar{\psi}} \right),$$

[9] A derivation is available upon request.

where

$$\mathbf{\Omega}_{\bar{\psi}} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[T^{-2} \sum_{t,s=1}^{T} \mathbf{d}_{r,it}\mathbf{d}'_{r,is} E\left(\varepsilon_{it}\varepsilon_{is}\right) \right], \tag{49}$$

where $\mathbf{d}_{r,it} = (\mathbf{r}_i - \bar{\mathbf{r}}) - \frac{1}{N}\sum_{j=1}^{N}(\mathbf{r}_j - \bar{\mathbf{r}})w_{ji,t}$, and $w_{ij,t} = (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \mathbf{Q}_{FE,NT}^{-1}(\mathbf{x}_{jt} - \bar{\mathbf{x}}_j)$, as before. Moreover, we note that under Assumption P9

$$\mathbf{Q}_{zr,N}\mathbf{Q}_{rr,N}^{-1}\mathbf{Q}'_{zr,N} \to_p \mathbf{Q}_{zr}\mathbf{Q}_{rr}^{-1}\mathbf{Q}'_{zr}, \text{ and } \mathbf{Q}_{zr,N}\mathbf{Q}_{rr,N}^{-1} \to_p \mathbf{Q}_{zr}\mathbf{Q}_{rr}^{-1}.$$

Using the above results and Slutsky's theorem now yields

$$\sqrt{N}\left(\hat{\boldsymbol{\gamma}}_{FEF-IV} - \boldsymbol{\gamma}\right) \to_d N\left(\mathbf{0}, \mathbf{\Omega}_{\hat{\gamma}_{FEF-IV}}\right),$$

where

$$\mathbf{\Omega}_{\hat{\gamma}_{FEF-IV}} = \left(\mathbf{Q}_{zr}\mathbf{Q}_{rr}^{-1}\mathbf{Q}'_{zr}\right)^{-1}\mathbf{Q}_{zr}\mathbf{Q}_{rr}^{-1}\left(\sigma_\eta^2\mathbf{Q}_{rr} + \Omega_{\bar{\psi}}\right)\mathbf{Q}_{rr}^{-1}\mathbf{Q}'_{zr}\left(\mathbf{Q}_{zr}\mathbf{Q}_{rr}^{-1}\mathbf{Q}'_{zr}\right)^{-1}. \tag{50}$$

For convenience, the above results are summarized in the following theorem.

Theorem 2 *Suppose that Assumptions P1-P6, P8 and P9 hold, and let the FEF-IV estimator be defined as in (48). Then we have*

$$\sqrt{N}\left(\hat{\boldsymbol{\gamma}}_{FEF-IV} - \boldsymbol{\gamma}\right) \to_d N\left(\mathbf{0}, \mathbf{\Omega}_{\hat{\gamma}_{FEF-IV}}\right),$$

where $\mathbf{\Omega}_{\hat{\gamma}_{FEF-IV}}$ *is given by (50).*

The variance of $\hat{\boldsymbol{\gamma}}_{FEF-IV}$ can now be estimated along similar lines as in Section 3.3. We have

$$\widehat{Var}\left(\hat{\boldsymbol{\gamma}}_{FEF-IV}\right) = N^{-1}\mathbf{H}_{zr,N}\left[\hat{\mathbf{V}}_{rr,N} + \mathbf{Q}_{r\bar{x},N}\left(N\widehat{Var}(\hat{\boldsymbol{\beta}})\right)\mathbf{Q}'_{r\bar{x},N}\right]\mathbf{H}'_{zr,N}, \tag{51}$$

where

$$\mathbf{H}_{zr,N} = \left(\mathbf{Q}_{zr,N}\mathbf{Q}_{rr,N}^{-1}\mathbf{Q}'_{zr,N}\right)^{-1}\mathbf{Q}_{zr,N}\mathbf{Q}_{rr,N}^{-1},$$

$$\mathbf{Q}_{r\bar{x},N} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{r}_i - \bar{\mathbf{r}}\right)\left(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}\right)',$$

$$\hat{\mathbf{V}}_{rr,N} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i - \bar{\mathbf{r}})(\mathbf{r}_i - \bar{\mathbf{r}})'\left(\hat{\upsilon}_i - \bar{\hat{\upsilon}}\right)^2,$$

$$\hat{\upsilon}_i - \bar{\hat{\upsilon}} = \bar{y}_i - \bar{y} - \left(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}\right)'\hat{\boldsymbol{\beta}} - (\mathbf{z}_i - \bar{\mathbf{z}})'\hat{\boldsymbol{\gamma}}_{FEF-IV},$$

and as before $\widehat{Var}(\hat{\boldsymbol{\beta}})$ is given by (18). It can be shown that $\widehat{Var}\left(\sqrt{N}\hat{\boldsymbol{\gamma}}_{FEF-IV}\right)$ is a consistent estimator of $\Omega_{\hat{\gamma}_{FEF-IV}}$ defined by (50) if condition (16) is met. Our simulation results suggest that the above variance estimator performs well even if condition (16) is not satisfied.

The main advantage of FEF-IV estimation procedure over the modified HT approach lies in the fact that it allows for errors to be heteroskedastic and serially correlated, and is relatively simple compute. But FEF-IV is likely to be less efficient than the HTM estimator when the errors are homoskedastic and serially uncorrelated.[10] Furthermore, in the exactly identified case, where $\dim(\mathbf{r}_i) = \dim(\mathbf{z}_{2,i})$, then the HTM estimator of γ, $\tilde{\boldsymbol{\gamma}}_{HT}$ defined by (44), is identical to the the FEF-IV estimator defined by (48).

Turning to the variance-covariance estimators of HTM and FEF-IV estimators, it is interesting to note that in general the variance-covariance matrix of the two estimators differ even if assume homoskedastic and serially uncorrelated errors. But as shown in a Supplement to the paper, the two estimators have the same variances if $\boldsymbol{\gamma}$ is exactly identified, $\bar{\mathbf{x}}_i$ and $\mathbf{r}_i$ are uncorrelated, and $\bar{\mathbf{z}} = \mathbf{0}, \bar{\mathbf{x}} = \mathbf{0}$. But as can be seen from the simulations, it is interesting that both estimators behave very similarly even if these conditions are not met.

In an empirical context, the differences between HT and FEF-IV estimators are illustrated in an application to return to schooling in Example 59 of Pesaran (2015). In this application a wage equation is estimated on a panel of $N = 545$ individuals from National Longitudinal Surveys of full-time working males over the period 1980 to 1987. The equation includes five time-varying regressors, namely experience, experience squared, three dummy variables for the marriage status, whether individual's wage is set by a union contract, and location (urban/rural), and three time-invariant regressors, namely education, and two dummy variables for black and Hispanic. This data set was originally analyzed by Vella and Verbeek (1998). The focus of the empirical analysis is on the effect of the education variable which is taken to be endogenously determined and is time-invariant in the sample. Very different results are obtained depending on which variables are taken to be exogenous. When we compute HT estimates using marriage status, Black and Hispanic as exogenous (one from the set of time-varying regressors and two from the set of time-invariant regressors), the coefficient of the education dummy is statistically insignificant and quite large. However, if FEF-IV estimation is used with the race variable, Hispanic, is treated as exogenous, then we find the education variable to be highly significant with a sensible magnitude (7%). Also in this specification, the race variable Black is statistically significant with negative effects on wages. Note that for this specification FEF-IV and HTM estimates will be the same since the time-invariant effects are exactly identified (see Section A of the Supplement).

5 Monte Carlo Simulation

In order to evaluate the performance of the FEF and the FEF-IV estimators proposed in this paper, we conducted three sets of simulations. One set with exogenous time-invariant regressors and the other sets where one of the time-invariant regressors is correlated with the fixed effects. In both sets of experiments the data generating process (DGP) include two time-varying and two time-invariant

[10] When the errors are homoskedastic and serially uncorrelated, the HTM estimator being a GMM type estimator is more efficient than the FEF-IV estimator which is a standard IV estimator.

regressors, and allow for error heteroskedasticity and residual serial correlation. Specifically, y_{it} is generated as

$$\begin{aligned} y_{it} &= 1 + \alpha_i + \beta_1 x_{it,1} + \beta_2 x_{it,2} + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \varepsilon_{it}, \\ i &= 1, 2, \ldots, N,\ t = 1, 2, \ldots, T, \end{aligned}$$

with $\beta_1 = \beta_2 = 1$ and $\gamma_1 = \gamma_2 = 1$. An intercept is included, and hence without loss of generality we generate the regressors with zero means. For the time-varying regressors we consider the following relatively general specifications

$$\begin{aligned} x_{it,1} &= \alpha_i g_{1t} + w_{it,1}, \\ x_{it,2} &= \alpha_i g_{2t} + w_{it,2}, \end{aligned}$$

where the time effects g_{1t} and g_{2t} are generated as $U(1,2)$ and are then kept fixed across the replications. Note that

$$\bar{x}_{i,j} = \alpha_i \bar{g}_j + \bar{w}_{i,j}, \text{ for } j = 1, 2,$$

where $\bar{x}_{i,j} = T^{-1} \sum_{t=1}^{T} x_{it,j}$, $\bar{w}_{ij} = T^{-1} \sum_{t=1}^{T} w_{it,j}$, and $\bar{g}_j = T^{-1} \sum_{t=1}^{T} g_{jt}$. We generate the fixed effects as $\alpha_i \sim 0.5\left(\chi^2(2) - 2\right)$, for $i = 1, 2, \ldots, N$. The stochastic components of the time varying regressors ($w_{it,1}$ and $w_{it,2}$) are generated as heterogenous stationary $AR(1)$ processes

$$w_{it,j} = \mu_{ij}(1 - \rho_{w,ij}) + \rho_{\omega,ij} w_{it-1,j} + \sqrt{1 - \rho_{w,ij}^2}\,\epsilon_{w,it,j} \text{ for } j = 1, 2,$$

where

$$\begin{aligned} \epsilon_{w,it,j} &\sim IIDN(0, \sigma_{\epsilon i}^2), \text{ for all } i, j \text{ and } t, \\ \sigma_{\epsilon i}^2 &\sim 0.5\left[1 + 0.5 IID\chi^2(2)\right],\ w_{i0,j} \sim IIDN(\mu_i, \sigma_{\epsilon i}^2), \text{ for all } i, j, \\ \rho_{\omega,ij} &\sim IIDU[0, 0.98],\ \mu_{ij} \sim IIDN(0, \sigma_\mu^2), \sigma_\mu^2 = 2, \text{for all } i, j. \end{aligned}$$

The above DGP allows the individual-specific means, $\bar{x}_{i,j}$, to be non-zero. Note also that the time-varying regressors, $x_{it,j}$, are correlated with the individual effects, α_i, since $\bar{g}_j \neq 0$, for $j = 1$ and 2.

The time-invariant regressors, z_{ji}, for $j = 1, 2$, are generated as

$$\mathbf{z}_i = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \boldsymbol{\Lambda} \bar{\mathbf{w}}_i + \alpha_i \boldsymbol{\phi} + \boldsymbol{\zeta}_i,$$

where $\mathbf{z}_i = (z_{i1}, z_{i2})'$, $\bar{\mathbf{w}}_i = (\bar{w}_{i1}, \bar{w}_{i2})'$, and $\boldsymbol{\zeta}_i \sim IIDN(\mathbf{0}, \mathbf{I}_2)$, and

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

We consider both cases where the time-invariant regressors, $\mathbf{z}_i$, are exogenous (DGP A) and when they are endogenous (DGP B and C). Under DGP A we set $\boldsymbol{\phi} = (\phi_1, \phi_2)' = \mathbf{0}$, and under DGP B and C we set $\boldsymbol{\phi} = (\phi_1, \phi_2)' = (1, 1)'$. We also assume the 4×1 vector of instruments $\mathbf{r}_i = (r_{i1}, r_{i2}, r_{i3}, r_{i4})'$ are generated as

$$\mathbf{r}_i = \boldsymbol{\Gamma}_{\zeta}\boldsymbol{\zeta}_i + \boldsymbol{\Gamma}_w \bar{\mathbf{w}}_i + \boldsymbol{\xi}_i, \tag{52}$$

where $\boldsymbol{\xi}_i \sim IIDN\left(\mathbf{0}, \mathbf{I}_4\right)$, with

$$\boldsymbol{\Gamma}_{\zeta} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix},$$

and distinguish between DGP B and C by different choices of $\boldsymbol{\Gamma}_w$, namely under DGP B we set $\boldsymbol{\Gamma}_w = 0$, and under DGP C we set

$$\boldsymbol{\Gamma}_w = 10 \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Clearly, DGP B and C are relevant only when the time-invariant regressors, $\mathbf{z}_i$, are endogenous. Under DGP B, $Cov(\mathbf{r}_i, \bar{\mathbf{x}}_i) = \mathbf{Q}_{r\bar{x}} = \mathbf{0}$, and under DGP C we have $\mathbf{Q}_{r\bar{x}} = \boldsymbol{\Gamma}_w \neq \mathbf{0}$. Given our theoretical derivations we expect the modified HT and FEF-IV estimators to perform very similarly under DGP B.

For each of the above three DGPs (A, B and C) we consider three different processes for the idiosyncratic errors, ε_{it}:

Case 1: Homoskedastic errors:

$$\varepsilon_{it} \sim IIDN(0, 1), \text{ for } i = 1, 2, \ldots, N; t = 1, 2, \ldots, T.$$

Case 2: Heteroscedastic errors:

$$\varepsilon_{it} \sim IIDN(0, \sigma_i^2), \; i = 1, 2, \ldots, N; t = 1, 2, \ldots, T,$$

where $\sigma_i^2 \sim 0.5\left[1 + 0.5IID\chi^2(2)\right]$ for all i.

Case 3: Serially correlated and heteroscedastic errors:

$$\varepsilon_{it} = \rho_{\varepsilon i}\varepsilon_{i,t-1} + \sqrt{1 - \rho_{\varepsilon i}^2}v_{it},$$

where

$$\begin{aligned} \varepsilon_{i0} &= 0 \text{ for all } i, \\ v_{it} &\sim IIDN(0, \sigma_{vi}^2), \text{ for all } i \text{ and } t, \\ \sigma_{vi}^2 &\sim 0.5(1 + 0.5IID\chi^2(2)), \\ \rho_{\varepsilon i} &\sim IIDU[0, 0.98], \text{for all } i, \end{aligned}$$

for $t = -49, -48, ..., 0, 1, 2, ..., T$, with $u_{i,-49} = 0$, for all i. The first 50 observations are discarded, and the remaining T observations are used in the experiments. We consider the simulation of combinations of $N = 500, 1000, 2000$ and $T = 3, 5, 10$.

Overall, we conduct 9 experiments, which we denote by $A(1)$, $A(2), A(3)$, $B(1), B(2), B(3)$, $C(1), C(2)$ and $C(3)$. The number in parentheses refers to the type of DGP used for the errors, ε_{it}, as specified under the cases (1)-(3). Experiments $A(i)$, for $i = 1, 2$, and 3, generates $\mathbf{z}_i$ as exogenous, whilst experiments $B(i)$ and $C(i)$ allow $\mathbf{z}_i$ to be endogenous, and considers the over-identified case where the number of instruments (4), is larger than the number of endogenous regressors. To save space we have not considered the exact identified case, where $\mathbf{r}_i$ has the same dimension as $\mathbf{z}_i$.

Throughout we carry out $1,000$ replications for each experiment, and report bias, root mean squared error (RMSE), size and power for different estimators of $\boldsymbol{\gamma}$, namely FEVD with and without intercepts in the second step, and the FEF estimator proposed in this paper for DGP A. We also consider HT and FEF-IV estimators in the case of DGP B and C where the time varying regressors are correlated with the errors. But to save space, in what follows we report only the simulation results for γ_1 in the case of DGP A and DGP C. The results for γ_2 are very similar to γ_1 and the results for DGP B are also similar to DGP C. A complete set of simulation results is available in the Supplement to the paper.

The results of FEF for DGP A are summarized in Tables 1-3, and clearly show that the FEF estimator performs well in all experiments, even when the errors are serially correlated and/or heteroskedastic. It has much lower bias and RMSE as compared to the FEVD estimator proposed by PT. However, in accordance with our theoretical findings, the FEF and FEVD estimators become identical when an intercept is included in the second stage of the PT estimation procedure. However, even after this correction the FEVD approach continues to exhibit substantial size distortions due to the use of incorrect standard errors in the third stage of the procedure (see Section 3.4).

The results for DGP C are summarized in Tables 4-6. In the case of these DGPs, the FEF-IV estimator is computed using $\mathbf{r}_i$ (defined by (52)) as an instrument for the endogenous time-invariant regressor, $\mathbf{z}_i$. The HTM estimator is the modified HT estimation uses $\mathbf{r}_i$ as the instruments (see (45)). The FEF-IV procedure performs well in all cases, irrespective of whether the errors are heteroskedastic and/or serially correlated. In particular the size of the FEF-IV estimator is very close to the 5% nominal value, with the power rising steadily in N. This suggests that the variance estimator for the FEF-IV, (51), is valid in the case of error heteroskedasticity and performs well even the errors are serially correlated (for example, see Tables 5 and 6). Similar results are also obtained for the modified HT estimator. Interestingly, in the case of theses experiments FEF-IV and the modified HT estimators perform similarly, suggesting that the third step of the HT estimator does not lead to efficiency gains over the two-step FEF-IV method.

Finally, considering that both time-varying regressors are correlated with the individual effects, the original (unmodified) version of HT estimator will fail to provide consistent estimators since none of the time-varying variables can be used as instruments for estimation of the time-invariant effects. The small sample properties of the HT estimator are investigated using additional Monte Carlo experiments. Not surprising we find the HT estimator to be biased with substantial size distortions. The results of these experiments are provided in the Supplement to the paper.

6 Conclusion

In this paper, we propose the FEF and FEF-IV estimators for panel models with time-invariant regressors. The FEF estimator is computed using a two-step procedure, where in the first step the fixed effects estimators are used to filter the effects of time-varying regressors. In the second step, time averages of the residuals are used in cross-section regressions to estimate the coefficients of time-invariant regressors. We also develop the asymptotic distribution for the FEF, and show that it's unbiased, consistent and asymptotically normally distributed. The FEF estimator is sufficiently robust and allows for cross-sectional heteroskedasticity and serial correlation. An alternative variance estimators of the FEF estimator is also proposed in this paper.

Moreover, when there is correlation between the time-invariant variables and individual effects, we propose the FEF-IV estimator, which can also be calculated by a two step procedure. The first step of FEF-IV is similar to FEF, but in the second step, we use the instrument variable estimation for the time-invariant regressors. We also show that this FEF-IV estimator is consistent and asymptotically normally distributed. Similar to the FEF estimator, the FEF-IV estimator is sufficiently robust and allows for cross-sectional heteroskedasticity and serial correlation. An alternative variance estimator of the FEF-IV estimator is also proposed in this paper. By simulations, we find both the FEF and FEF-IV have better small sample performance in terms of bias and RMSE, and most importantly has the correct size in the presence of correlation of arbitrary degree between the time-varying regressors and the individual effects.

Furthermore, we also contribute to the debate on the FEVD estimator proposed by Plumper and Troeger (2007). We show that the FEVD estimator is exactly the same as our FEF estimator if an intercept is included in the second step of PT's procedure, but the FEVD estimator is inconsistent in general if no intercept is included in the second stage (see equation (5) in PT). Furthermore, even if the FEVD estimator is computed using an intercept in the second stage, it will still lead to misleading inference since contrary to what is claimed by PT, the standard errors computed in the third stage of the PT procedure are not valid.

Overall, our Monte Carlo simulations show that FEF and FEF-IV estimators proposed in this paper perform well in terms of bias, RMSE, size and power. The simulation results also confirm our theoretical derivations showing that in general the FEVD estimator suffers from size distortions. Finally, in cases where none of the time-varying regressors is uncorrelated with the fixed effects, the use of standard HT procedure can lead to biased estimates and misleading inference. In such cases a modified version of the HT procedure is proposed which can be used if there exists a sufficient number of time-invariant instruments for the endogenous time-invariant regressors. We find that the FEF-IV and modified HT have very similar small sample properties and suggests that the third stage of the modified HT estimator is redundant and the simpler two-step FEF-IV estimator performs equally well. It is also illustrated that the original HT estimator can be badly biased worth substantial size distortions if all the time-varying regressors are correlated with the errors, a possibility that can not be ruled out in practice.

Appendix: Mathematical Derivations

Lemma A.1 *Suppose that* $\mathbf{A}$ *is a* $p \times p$ *symmetric matrix,* p *is fixed,* $\lambda_{\min}(\mathbf{A}) \geq 2/K$*, and* $\lambda_{\max}(\mathbf{A}) \leq K/2$*, with* K *being a fixed, non-zero positive constant. Consider now the stochastic matrix* $\hat{\mathbf{A}}_N$*, viewed as an estimator of* $\mathbf{A}$*, such that* $\left\|\hat{\mathbf{A}}_N - \mathbf{A}_N\right\| \rightarrow_p 0$*. Then with probability approaching one* $\lambda_{\min}\left(\hat{\mathbf{A}}_N\right) \geq 1/K$ *and* $\lambda_{\max}\left(\hat{\mathbf{A}}_N\right) \leq K$*.*

Source: Lemma A0 in the mathematical supplement to Newey and Windmeijer (2009).

Lemma A.2 *Given the cross-product sample moments defined by (26), (27) and (28), and in view of (23) and (24), we have*

$$\mathbf{Q}_{h\bar{y},N} = \mathbf{Q}_{h\bar{x},N}\hat{\boldsymbol{\beta}} + \mathbf{Q}_{hh,N}, \tag{A.1}$$

$$\mathbf{q}_{p,NT} = \mathbf{Q}_{p,NT}\hat{\boldsymbol{\beta}} + \mathbf{Q}'_{z\bar{x},N}\hat{\boldsymbol{\gamma}} + \mathbf{Q}'_{h\bar{x},N}, \tag{A.2}$$

$$\mathbf{Q}_{z\bar{y},N} = \mathbf{Q}_{z\bar{x},N}\hat{\boldsymbol{\beta}} + \mathbf{Q}_{zz,N}\hat{\boldsymbol{\gamma}}, \tag{A.3}$$

$$\mathbf{Q}_{zz,N}\left(\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}\right) + \mathbf{Q}_{z\bar{x},N}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right) = \mathbf{0}, \tag{A.4}$$

$$(\tilde{\delta} - 1)\mathbf{Q}_{hh,N} = \mathbf{Q}_{h\bar{x},N}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right). \tag{A.5}$$

Proof. Using (23) and (24) we first note that

$$\begin{aligned}
\mathbf{Q}_{h\bar{y},N} &= \frac{1}{N}\sum_{i=1}^{N}\left(\hat{h}_i - \bar{\hat{h}}\right)\left(\bar{y}_i - \bar{y}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\hat{h}_i - \bar{\hat{h}}\right)\left(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}\right)'\hat{\boldsymbol{\beta}} + \frac{1}{N}\sum_{i=1}^{N}\left(\hat{h}_i - \bar{\hat{h}}\right)\left(\mathbf{z}_i - \bar{\mathbf{z}}\right)'\hat{\boldsymbol{\gamma}} + \frac{1}{N}\sum_{i=1}^{N}\left(\hat{h}_i - \bar{\hat{h}}\right)^2 \\
&= \mathbf{Q}_{h\bar{x},N}\hat{\boldsymbol{\beta}} + \mathbf{Q}_{hh,N}.
\end{aligned}$$

Similarly

$$\begin{aligned}
\mathbf{q}_{p,NT} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}\right)\left(y_{it} - \bar{y}\right) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}\right)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}\right)'\hat{\boldsymbol{\beta}} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}\right)\left(\hat{u}_{it} - \bar{\hat{u}}\right) \\
&= \mathbf{Q}_{p,NT}\hat{\boldsymbol{\beta}} + \mathbf{Q}'_{z\bar{x},N}\hat{\boldsymbol{\gamma}} + \mathbf{Q}'_{h\bar{x},N},
\end{aligned}$$

where

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}\right)\left(\hat{u}_{it}-\bar{\hat{u}}\right) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\hat{u}_{it}-\bar{\hat{u}}_i\right)+\frac{1}{N}\sum_{i=1}^{N}\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)\left(\bar{\hat{u}}_i-\bar{\hat{u}}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)\left(\bar{\hat{u}}_i-\bar{\hat{u}}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\hat{\boldsymbol{\gamma}}+\frac{1}{N}\sum_{i=1}^{N}\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)\left(\hat{h}_i-\bar{\hat{h}}\right) \\
&= \mathbf{Q}'_{z\bar{x},N}\hat{\boldsymbol{\gamma}}+\mathbf{Q}'_{h\bar{x},N},
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{Q}_{z\bar{y},N} &= \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\bar{y}_i-\bar{y}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)'\hat{\boldsymbol{\beta}}+\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\bar{\hat{u}}_i-\bar{\hat{u}}\right) \\
&= \mathbf{Q}_{z\bar{x},N}\hat{\boldsymbol{\beta}}+\mathbf{Q}_{zz,N}\hat{\boldsymbol{\gamma}}.
\end{aligned}$$

Using this result together with (30) now yields

$$\mathbf{Q}_{zz,N}\left(\hat{\boldsymbol{\gamma}}-\tilde{\boldsymbol{\gamma}}\right)+\mathbf{Q}_{z\bar{x},N}\left(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}}\right)=\mathbf{0}.$$

Finally, (A.5) follows immediately using (A.1) and (31). ■

A.1 Proof of Theorem (1)

To derive the asymptotics of $\hat{\boldsymbol{\gamma}}_{FEF}$, we first note that the FE estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}=\left[\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)'\right]^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(y_{it}-\bar{y}_i\right). \tag{A.6}$$

Under the above assumptions, $\hat{\boldsymbol{\beta}}$ is unbiased and consistent for any fixed T and as $N\to\infty$, and

$$Var\left(\hat{\boldsymbol{\beta}}\,|\mathbf{X}\right)=\frac{1}{NT}\mathbf{Q}_{FE,NT}^{-1}\mathbf{V}_{FE,NT}\mathbf{Q}_{FE,NT}^{-1}, \tag{A.7}$$

where

$$\mathbf{V}_{FE,NT}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sigma_i^2\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)'+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t\neq s}^{T}\gamma_i(t,s)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{is}-\bar{\mathbf{x}}_i\right)'. \tag{A.8}$$

In the standard case where $\varepsilon_{it}\sim IID(0,\sigma^2)$, we obtain the more familiar expression $Var\left(\hat{\boldsymbol{\beta}}\,|\mathbf{X}\right)=(NT)^{-1}\sigma^2\mathbf{Q}_{FE,NT}^{-1}$. Also, for a fixed T and as $N\to\infty$, we have the following limiting distribution

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\to_d N(\mathbf{0},T^{-1}\boldsymbol{\Omega}_{\hat{\beta}}), \tag{A.9}$$

where

$$\boldsymbol{\Omega}_{\hat{\beta}}=\mathbf{Q}_{FE,T}^{-1}\mathbf{V}_{FE,T}\mathbf{Q}_{FE,T}^{-1}, \tag{A.10}$$

and $\mathbf{Q}_{FE,T}$ is defined in Assumption P5, and $\mathbf{V}_{FE,T}=p\lim_{N\to\infty}\left(\mathbf{V}_{FE,NT}\right)$.

Consider now the FEF estimator of $\boldsymbol{\gamma}$ defined by (4) and note that

$$\bar{\hat{u}}_i-\bar{\hat{u}}=\left(\eta_i-\bar{\eta}\right)+\left(\bar{\varepsilon}_i-\bar{\varepsilon}\right)+\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\boldsymbol{\gamma}-\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right).$$

Using this result in (4) we now have (noting that $N^{-1}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\bar{\varepsilon}+\bar{\eta})=\mathbf{0}$)

$$\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma}=\mathbf{Q}_{zz,N}^{-1}\left[N^{-1}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})\zeta_i\right], \tag{A.11}$$

where $\mathbf{Q}_{zz,N}$ is defined by (8) and

$$\zeta_i=\eta_i+\bar{\varepsilon}_i-(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right). \tag{A.12}$$

Let $\mathbf{Z}=(\mathbf{z}_1,....,\mathbf{z}_N)'$, $\mathbf{X}=(\mathbf{x}_{it};i=1,2,...,N;t=1,2,...,T)$, and $\boldsymbol{\eta}=(\eta_1,\eta_2,...,\eta_N)'$, and note that

$$\begin{aligned}E\left[(\mathbf{z}_i-\bar{\mathbf{z}})\zeta_i\,|\mathbf{Z},\mathbf{X},\boldsymbol{\eta}\right] &= (\mathbf{z}_i-\bar{\mathbf{z}})\eta_i-(\mathbf{z}_i-\bar{\mathbf{z}})(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\left[E\left(\hat{\boldsymbol{\beta}}\,|\mathbf{X}\right)-\boldsymbol{\beta}\right]\\ &= (\mathbf{z}_i-\bar{\mathbf{z}})\eta_i.\end{aligned}$$

Also under Assumption P7 we have $E\left[(\mathbf{z}_i-\bar{\mathbf{z}})\eta_i\right]=\mathbf{0}$ for all i, and using (A.11) it follows that $E\left(\hat{\boldsymbol{\gamma}}_{FEF}\right)=\boldsymbol{\gamma}$, which establishes that $\hat{\boldsymbol{\gamma}}_{FEF}$ is an unbiased estimator of $\boldsymbol{\gamma}$.

Consider now the consistency and the asymptotic distribution of $\hat{\boldsymbol{\gamma}}_{FEF}$. To this end we first note that

$$\begin{aligned}N^{-1}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})\zeta_i &= N^{-1}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\eta_i+\bar{\varepsilon}_i)\\ &\quad-\left[N^{-1}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\right]\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right).\end{aligned}$$

Also under Assumptions P6 and P7, $N^{-1}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\to_p\mathbf{Q}_{z\bar{x}}$,

$$N^{-1/2}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\eta_i+\bar{\varepsilon}_i)\to_d N(\mathbf{0},\omega_{iT}^2\mathbf{Q}_{zz}), \tag{A.13}$$

where

$$\omega_{iT}^2=\sigma_\eta^2+\frac{\sigma_i^2}{T}+\frac{1}{T^2}\sum_{s\neq t}\gamma_i(s,t) \tag{A.14}$$

with $\gamma_i(s,t)=E(\varepsilon_{is}\varepsilon_{it})$, and

$$N^{-1}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})\zeta_i=O_p\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)+O_p\left(N^{-1/2}\right).$$

Now using (A.11), and since $\mathbf{Q}_{zz,N}\to_p\mathbf{Q}_{zz}$, which is a non-singular matrix, then we also have

$$\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma}=O_p\left(N^{-1/2}\right)+O_p\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right). \tag{A.15}$$

Therefore, in view of (A.15) we obtain

$$\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma}=O_p\left(N^{-1/2}\right),$$

which establishes that $\hat{\boldsymbol{\gamma}}_{FEF}$, is a $\sqrt{N}$ consistent estimator of $\boldsymbol{\gamma}$.

Now let's turn to the asymptotic distribution of $\hat{\boldsymbol{\gamma}}_{FEF}$, we first note that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})\zeta_i=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})\eta_i+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})\left[\bar{\varepsilon}_i-(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right], \tag{A.16}$$

and consider the limiting distribution of the two terms of (A.16) and their covariance. We first note that the second term of the above can be written as

$$\begin{aligned} \bar{\varepsilon}_i - (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) &= \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it} - \frac{1}{T}\sum_{t=1}^{T} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \mathbf{Q}_{FE,NT}^{-1} \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N} (\mathbf{x}_{jt} - \bar{\mathbf{x}}_j)\,\varepsilon_{jt} \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_{it} - \frac{1}{N}\sum_{j=1}^{N} w_{ij,t}\varepsilon_{jt}\right), \end{aligned}$$

where[11]

$$w_{ij,t} = (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \mathbf{Q}_{FE,NT}^{-1} (\mathbf{x}_{jt} - \bar{\mathbf{x}}_j). \tag{A.17}$$

Hence,

$$\begin{aligned} & \frac{1}{\sqrt{N}}\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}}) \left[\bar{\varepsilon}_i - (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right] \\ = & \frac{1}{\sqrt{N}}\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}}) \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_{it} - \frac{1}{N}\sum_{j=1}^{N} w_{ij,t}\varepsilon_{jt}\right) \\ = & \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \bar{\boldsymbol{\xi}}_{i,N}, \end{aligned}$$

where

$$\bar{\boldsymbol{\xi}}_{i,N} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{d}_{z,it}\varepsilon_{it}, \tag{A.18}$$

and

$$\mathbf{d}_{z,it} = (\mathbf{z}_i - \bar{\mathbf{z}}) - \frac{1}{N}\sum_{j=1}^{N} (\mathbf{z}_j - \bar{\mathbf{z}})\, w_{ji,t}. \tag{A.19}$$

Using these results in (A.16) now yield

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}})\,\zeta_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}})\,\eta_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \bar{\boldsymbol{\xi}}_{i,N}. \tag{A.20}$$

However,

$$\begin{aligned} & Cov\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}})\,\eta_i, \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \bar{\boldsymbol{\xi}}_{i,N}\right) \\ = & \frac{1}{N}\sum_{i,j} E\left[(\mathbf{z}_i - \bar{\mathbf{z}})\,\eta_i \bar{\boldsymbol{\xi}}_{j,N}\right] = \frac{1}{N}\sum_{i,j} E\left\{E\left[\eta_i\,(\mathbf{z}_i - \bar{\mathbf{z}})\,\bar{\boldsymbol{\xi}}'_{j,N} | \mathbf{Z}, \mathbf{X}, \boldsymbol{\eta}\right]\right\} = \mathbf{0}, \end{aligned}$$

and it is sufficient to derive the asymptotic distributions of the two terms in (A.20), separately. To this end we note that under Assumptions P6 and P7, and using standard central limit theorems it follows that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}})\,\eta_i \to_d N\left(\mathbf{0}, \sigma_\eta^2 \mathbf{Q}_{zz}\right). \tag{A.21}$$

Consider now the second term in (A.20) and note that under Assumption P1-P3 and P7, $w_{ji,t}$ and $\mathbf{z}_i$ are distributed independently of ε_{is}, for all i, j, t, and s, and hence conditional on $\mathbf{Z}$ and $\mathbf{X}$, $\bar{\boldsymbol{\xi}}_{i,N}$ have zero means, and are cross sectionally independently distributed (noting that by Assumption P2, ε_{it} are assumed to be cross-sectionally independent). But since the terms, $\bar{\boldsymbol{\xi}}_{i,N}$, in (A.20) vary with N it suffices to show that the following Liapunov condition, (see Davidson (1994), p. 373) is

[11] Note that $w_{ij,t} \neq w_{ji,t}$.

satisfied.

$$\lim_{N\to\infty}\sum_{i=1}^{N}E\left\|N^{-1/2}\bar{\boldsymbol{\xi}}_{i,N}\right\|^{2+\delta}=0 \text{ for some } \delta>0. \tag{A.22}$$

The validity of this condition is established under Assumptions P1-P7 in Section A.2 of the Appendix. Hence

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{\boldsymbol{\xi}}_{i,N}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(T^{-1}\sum_{t=1}^{T}\mathbf{d}_{z,it}\varepsilon_{it}\right)\rightarrow_{d}N\left(\mathbf{0},\boldsymbol{\Omega}_{\bar{\xi}}\right) \tag{A.23}$$

where $\boldsymbol{\Omega}_{\bar{\xi}}=\lim_{N\to\infty}\boldsymbol{\Omega}_{\bar{\xi},N}$, and (since ε_{it} are assumed to be cross-sectionally independent)

$$\boldsymbol{\Omega}_{\bar{\xi},N}=N^{-1}\sum_{i=1}^{N}\left[T^{-2}\sum_{t,s=1}^{T}\mathbf{d}_{z,it}\mathbf{d}_{z,is}'E\left(\varepsilon_{it}\varepsilon_{is}\right)\right]. \tag{A.24}$$

A.2 Proof of Liapunov condition (A.22)

The Liapunov condition (A.22), for $\delta=2$ can be written as

$$\lim_{N\to\infty}N^{-2}\sum_{i=1}^{N}E\left\|\bar{\boldsymbol{\xi}}_{i,N}\right\|^{4}=0. \tag{A.25}$$

From (A.18) we first note that

$$\left\|\bar{\boldsymbol{\xi}}_{i,N}\right\|\leq\frac{1}{T}\sum_{t=1}^{T}\left\|\mathbf{d}_{z,it}\varepsilon_{it}\right\|,$$

and by Holder inequality

$$\sum_{t=1}^{T}\left\|\mathbf{d}_{z,it}\varepsilon_{it}\right\|\leq\left(\sum_{t=1}^{T}\left\|\mathbf{d}_{z,it}\right\|^{4}\right)^{1/4}\left(\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{4/3}\right)^{3/4},$$

and hence

$$\left\|\bar{\boldsymbol{\xi}}_{i,N}\right\|^{4}\leq\frac{1}{T^{4}}\left(\sum_{t=1}^{T}\left\|\mathbf{d}_{z,it}\right\|^{4}\right)\left(\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{4/3}\right)^{3}.$$

But under Assumptions P3 and P7, $\mathbf{x}_{it}$ and $\mathbf{z}_{i}$ are distributed independently of ε_{it}, and it follows that

$$E\left\|\bar{\boldsymbol{\xi}}_{i,N}\right\|^{4}\leq\left(T^{-1}\sum_{t=1}^{T}E\left\|\mathbf{d}_{z,it}\right\|^{4}\right)E\left[\left(T^{-1}\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{4/3}\right)^{3}\right].$$

But under Assumption P1, $E\left(\left|\varepsilon_{it}\right|^{4}\right)<K$, and since T is finite and for each i, ε_{it} are serially independent, then for some positive finite constant K_1 we have

$$E\left[\left(T^{-1}\sum_{t=1}^{T}\left|\varepsilon_{it}\right|^{4/3}\right)^{3}\right]\leq K_{1}\,T^{-1}\sum_{t=1}^{T}E\left|\varepsilon_{it}\right|^{4}<K_{1}K.$$

Hence,

$$N^{-2}\sum_{i=1}^{N}\left\|\bar{\boldsymbol{\xi}}_{i,N}\right\|^{4}<K_{1}KT^{-1}\sum_{t=1}^{T}\left[N^{-2}\sum_{i=1}^{N}E\left\|\mathbf{d}_{z,it}\right\|^{4}\right] \tag{A.26}$$

Now recall that

$$\mathbf{d}_{z,it}=\left(\mathbf{z}_{i}-\bar{\mathbf{z}}\right)-\frac{1}{N}\sum_{j=1}^{N}\left(\mathbf{z}_{j}-\bar{\mathbf{z}}\right)w_{ji,t},$$

where $w_{ji,t} = (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})' \mathbf{Q}_{FE,NT}^{-1} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$. Then

$$\begin{aligned} \mathbf{d}_{z,it} &= (\mathbf{z}_i - \bar{\mathbf{z}}) - \frac{1}{N}\sum_{j=1}^{N} (\mathbf{z}_j - \bar{\mathbf{z}}) (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})' \mathbf{Q}_{FE,NT}^{-1} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \\ &= \mathbf{z}_i - \bar{\mathbf{z}} - \mathbf{A}_{z\bar{x},N} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i), \end{aligned} \tag{A.27}$$

where $\mathbf{A}_{z\bar{x},N} = \mathbf{Q}_{z\bar{x},N}\mathbf{Q}_{FE,NT}^{-1}$. But[12]

$$\begin{aligned} \left\|\mathbf{A}_{z\bar{x},N}\right\|^2 &= tr\left(\mathbf{Q}_{FE,NT}^{-1}\mathbf{Q}_{z\bar{x},N}'\mathbf{Q}_{z\bar{x},N}\mathbf{Q}_{FE,NT}^{-1}\right) = tr\left(\mathbf{Q}_{FE,NT}^{-2}\mathbf{Q}_{z\bar{x},N}'\mathbf{Q}_{z\bar{x},N}\right) \\ &\leq \lambda_{\max}\left(\mathbf{Q}_{FE,NT}^{-2}\right) tr\left(\mathbf{Q}_{z\bar{x},N}'\mathbf{Q}_{z\bar{x},N}\right) \\ &= \lambda_{\max}^2\left(\mathbf{Q}_{FE,NT}^{-1}\right)\left\|\mathbf{Q}_{z\bar{x},N}\right\|^2 \\ &= \lambda_{\min}^{-2}\left(\mathbf{Q}_{FE,NT}\right)\left\|\mathbf{Q}_{z\bar{x},N}\right\|^2 \\ &= \frac{\left\|\mathbf{Q}_{z\bar{x},N}\right\|^2}{\lambda_{\min}^2\left(\mathbf{Q}_{FE,NT}\right)}, \end{aligned}$$

and noting that under Assumption P5, $\lambda_{\min}\left(\mathbf{Q}_{FE,NT}\right) > 1/K$, then

$$\left\|\mathbf{A}_{z\bar{x},N}\right\| \leq K\left\|\mathbf{Q}_{z\bar{x},N}\right\|. \tag{A.28}$$

Also, it is easily seen that

$$\left\|\mathbf{d}_{z,it}\right\|^2 = \mathbf{d}_{z,it}'\mathbf{d}_{z,it} = \left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^2 - 2(\mathbf{z}_i - \bar{\mathbf{z}})'\mathbf{A}_{z\bar{x},N}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\mathbf{A}_{z\bar{x},N}'\mathbf{A}_{z\bar{x},N}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i),$$

and

$$\begin{aligned} \left\|\mathbf{d}_{z,it}\right\|^4 &= \left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^4 + 4(\mathbf{z}_i - \bar{\mathbf{z}})'\mathbf{A}_{z\bar{x},N}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\mathbf{A}_{z\bar{x},N}'(\mathbf{z}_i - \bar{\mathbf{z}}) \\ &+ (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\mathbf{A}_{z\bar{x},N}'\mathbf{A}_{z\bar{x},N}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\mathbf{A}_{z\bar{x},N}'\mathbf{A}_{z\bar{x},N}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \\ &-4\left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^2(\mathbf{z}_i - \bar{\mathbf{z}})'\mathbf{A}_{z\bar{x},N}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + 2\left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^2(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\mathbf{A}_{z\bar{x},N}'\mathbf{A}_{z\bar{x},N}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \\ &-4(\mathbf{z}_i - \bar{\mathbf{z}})'\mathbf{A}_{z\bar{x},N}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\mathbf{A}_{z\bar{x},N}'\mathbf{A}_{z\bar{x},N}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i). \end{aligned}$$

Hence

$$\begin{aligned} \left\|\mathbf{d}_{z,it}\right\|^4 &\leq \left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^4 + 4\left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^2\left\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right\|^2\left\|\mathbf{A}_{z\bar{x},N}\right\|^2 \\ &+ \left\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right\|^4\left\|\mathbf{A}_{z\bar{x},N}\right\|^4 + 4\left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^3\left\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right\|\left\|\mathbf{A}_{z\bar{x},N}\right\| \\ &+2\left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^2\left\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right\|^2\left\|\mathbf{A}_{z\bar{x},N}\right\|^2 + 4\left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|\left\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right\|^3\left\|\mathbf{A}_{z\bar{x},N}\right\|^3, \end{aligned}$$

and using (A.28) we have

$$\begin{aligned} N^{-1}\sum_{i=1}^{N}\left\|\mathbf{d}_{z,it}\right\|^4 &\leq N^{-1}\sum_{i=1}^{N}\left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^4 + 4K_1^2\left\|\mathbf{Q}_{z\bar{x},N}\right\|^2\left[N^{-1}\sum_{i=1}^{N}\left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^2\left\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right\|^2\right] \\ &+K_1^4\left\|\mathbf{Q}_{z\bar{x},N}\right\|^4\left[N^{-1}\sum_{i=1}^{N}\left\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right\|^4\right] + 4K_1\left\|\mathbf{Q}_{z\bar{x},N}\right\|\left[N^{-1}\sum_{i=1}^{N}\left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|^3\left\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right\|\right] \\ &+4K_1^3\left\|\mathbf{Q}_{z\bar{x},N}\right\|^3\left[N^{-1}\sum_{i=1}^{N}\left\|\mathbf{z}_i - \bar{\mathbf{z}}\right\|\left\|\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right\|^3\right]. \end{aligned}$$

[12] Note that for any $p \times p$ matrices $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{A}$ is symmetric and $\mathbf{B}$ positive semi-definite, then $tr(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{A})\, tr(\mathbf{B})$.

Using this result in (A.26) we now obtain

$$N^{-1}\sum_{i=1}^{N}\left\|\bar{\boldsymbol{\xi}}_{i,N}\right\|^4 \leq K_1 K N^{-1}\sum_{i=1}^{N}\|\mathbf{z}_i-\bar{\mathbf{z}}\|^4+W_{1N}+W_{2N}+W_{3N}+W_{4N},$$

where

$$\begin{aligned}
W_{1N} &= 4KK_1^3\left\|\mathbf{Q}_{z\bar{x},N}\right\|^2\left[(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\mathbf{z}_i-\bar{\mathbf{z}}\|^2\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|^2\right],\\
W_{2N} &= KK_1^5\left\|\mathbf{Q}_{z\bar{x},N}\right\|^4\left[(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|^4\right],\\
W_{3N} &= 4KK_1^2\left\|\mathbf{Q}_{z\bar{x},N}\right\|\left[(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\mathbf{z}_i-\bar{\mathbf{z}}\|^3\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|\right],\\
W_{4N} &= 4KK_1^4\left\|\mathbf{Q}_{z\bar{x},N}\right\|^3\left[N^{-1}\sum_{i=1}^{N}\|\mathbf{z}_i-\bar{\mathbf{z}}\|\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|^3\right].
\end{aligned}$$

To investigate the limiting property of $N^{-1}\sum_{i=1}^{N}\left\|\bar{\boldsymbol{\xi}}_{i,N}\right\|^4$, we first note that by Assumption P6 $\left\|\mathbf{Q}_{z\bar{x},N}\right\|\to_p c$, as $N\to\infty$, where c is a finite constant, and by Slutsky's theorem (as $N\to\infty$, for a fixed T) we have

$$\begin{aligned}
W_{1N} &\to {}_p 4KK_1^3c^2\left[\lim_{N\to\infty}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left[\|\mathbf{z}_i-\bar{\mathbf{z}}\|^2\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|^2\right]\right],\\
W_{2N} &\to {}_p KK_1^5c^4\left[\lim_{N\to\infty}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|^4\right],\\
W_{3N} &\to {}_p 4KK_1^2c\left[\lim_{N\to\infty}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left[\|\mathbf{z}_i-\bar{\mathbf{z}}\|^3\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|\right]\right],\\
W_{4N} &\to {}_p 4KK_1^4c^3\left[\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\|\mathbf{z}_i-\bar{\mathbf{z}}\|\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|^3\right].
\end{aligned}$$

Hence

$$\begin{aligned}
\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}E\left\|\bar{\boldsymbol{\xi}}_{i,N}\right\|^4 \leq\; & K\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}E\|\mathbf{z}_i-\bar{\mathbf{z}}\|^4\\
&+4KK_1^3c^2\left[\lim_{N\to\infty}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left[\|\mathbf{z}_i-\bar{\mathbf{z}}\|^2\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|^2\right]\right]\\
&+KK_1^5c^4\left[\lim_{N\to\infty}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|^4\right]\\
&+4KK_1^2c\left[\lim_{N\to\infty}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left[\|\mathbf{z}_i-\bar{\mathbf{z}}\|^3\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|\right]\right]\\
&+4KK_1^4c^3\left[\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\|\mathbf{z}_i-\bar{\mathbf{z}}\|\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\|^3\right].
\end{aligned}$$

Therefore, $N^{-1}\sum_{i=1}^{N}E\left\|\bar{\boldsymbol{\xi}}_{i,N}\right\|^4$ is bounded and converges to a finite limit as $N\to\infty$ (irrespective of whether T is fixed or tends to infinity) if the following conditions hold

$$\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}E\|\mathbf{z}_i-\bar{\mathbf{z}}\|^4<K,$$

$$\begin{aligned}\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|^4 &< K, \text{ when } T \text{ is fixed} \\ \lim_{N \text{ and } T\to\infty} (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|^4 &< K\end{aligned}$$

$$\begin{aligned}\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E\left[\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^2\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|^2\right] &< K, \text{ when } T \text{ is fixed,} \\ \lim_{N \text{ and } T\to\infty} (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left[\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^2\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|^2\right] &< K,\end{aligned} \tag{A.29}$$

$$\begin{aligned}\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E\left[\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^3\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|\right] &< K, \text{ when } T \text{ is fixed,} \\ \lim_{N \text{ and } T\to\infty} (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left[\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^3\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|\right] &< K,\end{aligned} \tag{A.30}$$

$$\begin{aligned}\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E\left[\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|^3\right] &< K, \text{ when } T \text{ is fixed,} \\ \lim_{N \text{ and } T\to\infty} (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left[\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|^3\right] &< K,\end{aligned} \tag{A.31}$$

The above conditions are clearly satisfied if $\mathbf{x}_{it}$ and $\mathbf{z}_i$ have bounded supports. In the case where $\mathbf{x}_{it}$ and $\mathbf{z}_i$ do not have bounded supports the following moment conditions are sufficient to ensure that $\lim_{N\to\infty} N^{-2}\sum_{i=1}^{N} E\left\|\bar{\boldsymbol{\xi}}_{i,N}\right\|^4 = 0$, as required (applying Cauchy–Schwarz inequality to (A.29) and Holder's Inequality to (A.30) and (A.31)[13]

$$E\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^4 < K, \text{ and } E\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|^4 < K,$$

for all i,t, N and T. These conditions allow for any degree of dependence between $\mathbf{z}_i$ and $\mathbf{x}_{it}$.

A.3 Proof of Proposition 1

Using (A.27) and noting that $E(\varepsilon_{is}\varepsilon_{it}) = \gamma_i(s,t)$, $\mathbf{\Omega}_{\bar{\xi},N}$ defined by (A.24) can be written as

$$\begin{aligned}\mathbf{\Omega}_{\bar{\xi},N} &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t,s=1}^{T}\mathbf{d}_{z,it}\mathbf{d}'_{z,is}E(\varepsilon_{it}\varepsilon_{is}) \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t,s=1}^{T}\left[\mathbf{z}_i-\bar{\mathbf{z}}-\mathbf{A}_{z\bar{x}}(\mathbf{x}_{it}-\bar{\mathbf{x}}_i)\right]\left[\mathbf{z}_i-\bar{\mathbf{z}}-\mathbf{A}_{z\bar{x}}(\mathbf{x}_{is}-\bar{\mathbf{x}}_i)\right]'\gamma_i(t,s)\end{aligned}$$

[13] Note that by Holder's inequality

$$E\left[\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^3\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|\right] \le \left[E\left(\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^4\right)\right]^{3/4}\left[E\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|^4\right]^{1/4},$$

and

$$E\left[\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|^3\right] \le \left[E\left(\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^4\right)\right]^{1/4}\left[E\left\|\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right\|^4\right]^{3/4}.$$

where $\mathbf{A}_{z\bar{x},N}=\mathbf{Q}_{z\bar{x},N}\mathbf{Q}_{FE,NT}^{-1}$. Since

$$\begin{aligned} & \left[\mathbf{z}_i-\bar{\mathbf{z}}-\mathbf{A}_{z\bar{x},N}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\right]\left[\mathbf{z}_i-\bar{\mathbf{z}}-\mathbf{A}_{z\bar{x},N}\left(\mathbf{x}_{is}-\bar{\mathbf{x}}_i\right)\right]' \\ = & \left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'-\mathbf{A}_{z\bar{x},N}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'-\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{x}_{is}-\bar{\mathbf{x}}_i\right)'\mathbf{A}_{z\bar{x},N}' \\ & +\mathbf{A}_{z\bar{x},N}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{is}-\bar{\mathbf{x}}_i\right)'\mathbf{A}_{z\bar{x},N}', \end{aligned}$$

then setting $\kappa_{iT}^2=T^{-2}\sum_{t,s=1}^{T}\gamma_i(t,s)$, we obtain (See (A.14) and (13)).

$$\sigma_\eta^2\mathbf{Q}_{zz}+\boldsymbol{\Omega}_{\bar{\xi},N}=\mathring{\mathbf{Q}}_{zz,N}+\boldsymbol{\Delta}_N-\boldsymbol{\Delta}_{\bar{\xi}N}-\boldsymbol{\Delta}_{\bar{\xi}N}' \tag{A.32}$$

where

$$\mathring{\mathbf{Q}}_{zz,N}=\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_\eta^2+\kappa_{iT}^2\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)',$$

$$\boldsymbol{\Delta}_N=\frac{1}{T}\mathbf{A}_{z\bar{x},N}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t,s=1}^{T}\gamma_i(t,s)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{is}-\bar{\mathbf{x}}_i\right)'\right]\mathbf{A}_{z\bar{x},N}', \tag{A.33}$$

and

$$\boldsymbol{\Delta}_{\bar{\xi},N}=\mathbf{A}_{z\bar{x},N}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t,s=1}^{T}\gamma_i(t,s)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\right]. \tag{A.34}$$

Consider (A.33) and note that using (A.8) and replacing $\mathbf{A}_{z\bar{x},N}$ by $\mathbf{Q}_{z\bar{x},N}\mathbf{Q}_{FE,NT}^{-1}$, it can be written as

$$\boldsymbol{\Delta}_N=T^{-1}\mathbf{Q}_{z\bar{x},N}\mathbf{Q}_{FE,NT}^{-1}\mathbf{V}_{FE,NT}\mathbf{Q}_{FE,NT}^{-1}\mathbf{Q}_{z\bar{x},N}',$$

which upon using (A.7) reduces to

$$\boldsymbol{\Delta}_N=\mathbf{Q}_{z\bar{x},N}Var(\sqrt{N}\hat{\boldsymbol{\beta}})\mathbf{Q}_{z\bar{x},N}'. \tag{A.35}$$

Similarly, using the expression for $\mathbf{A}_{z\bar{x},N}$ given above, $\Delta_{\bar{\xi},N}$, given by (A.34) can be written as

$$\boldsymbol{\Delta}_{\bar{\xi},N}=\mathbf{Q}_{z\bar{x},N}\mathbf{Q}_{FE,NT}^{-1}\left[\frac{1}{T^2N}\sum_{i=1}^{N}\sum_{t,s=1}^{T}\gamma_i(t,s)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\right]. \tag{A.36}$$

Substituting (A.35) and (A.36) in (A.32) now yields the desired result.

A.4 Proof of Proposition 2

Consider (11) and the decomposition (14), and note that under Assumptions P1-P5, $\widehat{Var}(\sqrt{N}\hat{\boldsymbol{\beta}})$ defined by (18) tends to $Var\left(\sqrt{N}\hat{\boldsymbol{\beta}}\right)$ for a fixed T and as $N\to\infty$. (See, for example, Arellano (1987)). Consider now $\hat{\mathbf{V}}_{zz,N}$ defined by (19) and note that $\hat{\varsigma}_i-\bar{\hat{\varsigma}}$ defined by (20) can be written as

$$\hat{\varsigma}_i-\bar{\hat{\varsigma}}=(\eta_i-\bar{\eta})+(\bar{u}_i-\bar{u})-(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)-(\mathbf{z}_i-\bar{\mathbf{z}})'\left(\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma}\right).$$

Then

$$\hat{\mathbf{V}}_{zz,N}=\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\left(\hat{\varsigma}_i-\bar{\hat{\varsigma}}\right)^2=\mathbf{A}_{1N}+\mathbf{A}_{2N}-\mathbf{A}_{3N},$$

where

$$\begin{aligned}
\mathbf{A}_{1N} &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'\left[(\eta_i-\bar{\eta})+(\bar{\varepsilon}_i-\bar{\varepsilon})\right]^2,\\
\mathbf{A}_{2N} &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'\left[(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)+(\mathbf{z}_i-\bar{\mathbf{z}})'(\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma})\right]^2\\
\mathbf{A}_{3N} &= \frac{2}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'\left[(\eta_i-\bar{\eta})+(\bar{\varepsilon}_i-\bar{\varepsilon})\right]\left[(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)+(\mathbf{z}_i-\bar{\mathbf{z}})'(\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma})\right]
\end{aligned}$$

Starting with $\mathbf{A}_{1N}$, we have

$$\begin{aligned}
\mathbf{A}_{1N} &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'\left[v_i-(\bar{\eta}+\bar{\varepsilon})\right]^2\\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'v_i^2-2(\bar{\eta}+\bar{\varepsilon})\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'v_i\\
&\quad+\frac{(\bar{\eta}+\bar{\varepsilon})^2}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'.
\end{aligned}$$

Under Assumptions P1, P2 and P7, $\bar{\eta}+\bar{\varepsilon}=O_p\left(N^{-1/2}\right)+O_p\left(N^{-1/2}T^{-1/2}\right)=o_p(1)$. Also, $N^{-1}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'v_i=o_p(1)$, and conditional on $\mathbf{Z}=(\mathbf{z}_1,\mathbf{z}_2,...,\mathbf{z}_N)'$,

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'\left[v_i^2-E(v_i^2)\right]=o_p(1),$$

as $N\to\infty$. This latter result follows under Assumption P7, $\xi_i=v_i^2-E(v_i^2)$ and $\mathbf{z}_i$ are independently distributed, and v_i are cross-sectionally independent. Under these assumptions

$$E\left[\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'\left[v_i^2-E(v_i^2)\right]\right]=\frac{1}{N}\sum_{i=1}^{N}E\left[(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'\right]E\left[v_i^2-E(v_i^2)\right]=\mathbf{0},$$

and

$$E\left|\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'\left[v_i^2-E(v_i^2)\right]\right|\leq\frac{1}{N}\sum_{i=1}^{N}E\left\|(\mathbf{z}_i-\bar{\mathbf{z}})\right\|^2E\left|v_i^2-E(v_i^2)\right|<K<\infty,$$

since by assumption $E\left\|(\mathbf{z}_i-\bar{\mathbf{z}})\right\|^2<K$, and $E(v_i^2)<K$. In view of these results it now follows that

$$\mathbf{A}_{1N}=\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'E\left(v_i^2\right)+o_p(1).$$

Consider now $\mathbf{A}_{2N}$ and note similarly that

$$\begin{aligned}
\mathbf{A}_{2N} &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'\left[(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)+(\mathbf{z}_i-\bar{\mathbf{z}})'(\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma})\right]^2\\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})\\
&\quad+\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'(\mathbf{z}_i-\bar{\mathbf{z}})'(\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma})(\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma})'(\mathbf{z}_i-\bar{\mathbf{z}})\\
&\quad+\frac{2}{N}\sum_{i=1}^{N}(\mathbf{z}_i-\bar{\mathbf{z}})(\mathbf{z}_i-\bar{\mathbf{z}})'(\bar{\mathbf{x}}_i-\bar{\mathbf{x}})'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)(\mathbf{z}_i-\bar{\mathbf{z}})'(\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma}),
\end{aligned}$$

and expectations of all the three terms above tend to zero with N. Furthermore

$$\begin{aligned}&\left\| N^{-1}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\right\| \\ \leq\ &\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2\left[N^{-1}\sum_{i=1}^{N}\left\|\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)'\right\|^2\right].\end{aligned}$$

But by Cauchy–Schwarz inequality

$$N^{-1}\sum_{i=1}^{N}E\left[\left\|\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)'\right\|^2\right]\leq N^{-1}\sum_{i=1}^{N}\left(E\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^4\right)^{1/2}\left(E\left\|\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right\|^4\right)^{1/2},$$

and since under Assumption P4 and P7, $E\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^4<K$ and $E\left\|\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right\|^4<K$, it then follows that $N^{-1}\sum_{i=1}^{N}\left\|\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)'\right\|^2$ converges to a finite limit and hence

$$E\left\|N^{-1}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\right\|\leq K\ E\left[\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2\right]=O\left(N^{-1}\right).$$

A similar line of argument applies to other terms of $\mathbf{A}_{2,N}$.

Finally, for $\mathbf{A}_{3N}$ we have

$$\mathbf{A}_{3N}=\frac{2}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\left[v_i-\left(\bar{\eta}+\bar{\varepsilon}\right)\right]\left[\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)+\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'\left(\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma}\right)\right].$$

Once again noting that $\bar{\eta}+\bar{\varepsilon}=o_p(1)$, $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}=O_p\left(N^{-1/2}\right)=o_p\left(1\right)$, $\hat{\boldsymbol{\gamma}}_{FEF}-\boldsymbol{\gamma}=O_p\left(N^{-1/2}\right)$, it then follows that

$$\begin{aligned}\mathbf{A}_{3N}\ =\ &\left(\frac{2}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'v_i\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)'\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\\ &-\left(\frac{2}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'v_i(\mathbf{z}_i-\bar{\mathbf{z}})'\right)\mathbf{Q}_{zz,N}^{-1}\mathbf{Q}_{z\bar{x},N}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)+o_p(1)\end{aligned}$$

so long as

$$E\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'v_i\left(\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right)'\right\|\leq\frac{1}{N}\sum_{i=1}^{N}E\left[\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^2\left\|\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right\|\right]E\left|v_i\right|<K<\infty,$$

and

$$E\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'v_i(\mathbf{z}_i-\bar{\mathbf{z}})'\right\|\leq\frac{1}{N}\sum_{i=1}^{N}E\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^3E\left|v_i\right|<K<\infty.$$

The above conditions are met if $E\left\|\mathbf{z}_i-\bar{\mathbf{z}}\right\|^4<K$ and $E\left\|\bar{\mathbf{x}}_i-\bar{\mathbf{x}}\right\|^2<K$, for all i.

Considering all the three terms together we now have

$$\hat{\mathbf{V}}_{zz,N}=\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'E(v_i^2)+o_p(1).$$

We also note that $E(v_i^2)=\omega_{iT}^2$, where ω_{iT}^2 is defined by (A.14), and hence $\hat{\mathbf{V}}_{zz,N}\rightarrow_p\mathbf{V}_{zz}$, defined by (12) as required.

Finally, since $\mathbf{Q}_{z\bar{x},N}\mathbf{Q}_{FE,NT}^{-1}\rightarrow_p\mathbf{Q}_{z\bar{x}}\mathbf{Q}_{FE,T}^{-1}$, as $N\rightarrow\infty$, which is finite and bounded in N, then for a fixed T, $\Delta_{\bar{\xi},N}$ (defined by (15)) has the same order as $N^{-1}T^{-2}\sum_{i=1}^{N}\sum_{t,s=1}^{T}\gamma_i(t,s)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{z}_i-\bar{\mathbf{z}}\right)'$, and for a fixed T then $\boldsymbol{\Delta}_{\bar{\xi},N}=o_p(1)$, if condition (16) is met.

A.5 Proof of proposition 3

Rewrite the normal equations of the FEVD procedure, (29)-(31) in the following matrix format

$$\begin{pmatrix} \mathbf{Q}_{p,NT} & \mathbf{Q}'_{z\bar{x},N} & \mathbf{Q}'_{h\bar{x},N} \\ \mathbf{Q}_{z\bar{x},N} & \mathbf{Q}_{zz,N} & 0 \\ \mathbf{Q}_{h\bar{x},N} & 0 & \mathbf{Q}_{hh,N} \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\gamma}} \\ \tilde{\delta} \end{pmatrix} = \begin{pmatrix} \mathbf{q}_{p,NT} \\ \mathbf{Q}_{z\bar{y},N} \\ \mathbf{Q}_{h\bar{y},N} \end{pmatrix},$$

and note that the inverse of the LHS coefficient matrix is given by (see Magnus and Neudecker (2007), p12))

$$\begin{pmatrix} \mathbf{Q}_{p,NT} & \mathbf{Q}'_{z\bar{x},N} & \mathbf{Q}'_{h\bar{x},N} \\ \mathbf{Q}_{z\bar{x},N} & \mathbf{Q}_{zz,N} & 0 \\ \mathbf{Q}_{h\bar{x},N} & 0 & \mathbf{Q}_{hh,N} \end{pmatrix}^{-1} =$$

$$\begin{pmatrix} \mathbf{Q}_{NT}^{-1} & -\mathbf{Q}_{NT}^{-1}\mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1} & -\mathbf{Q}_{NT}^{-1}\mathbf{Q}'_{h\bar{x},N}\mathbf{Q}_{hh,N}^{-1} \\ -\mathbf{Q}_{zz,N}^{-1}\mathbf{Q}_{z\bar{x},N}\mathbf{Q}_{NT}^{-1} & \mathbf{Q}_{zz,N}^{-1} + \mathbf{Q}_{zz,N}^{-1}\mathbf{Q}_{z\bar{x},N}\mathbf{Q}_{NT}^{-1}\mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1} & \mathbf{Q}_{zz,N}^{-1}\mathbf{Q}_{z\bar{x},NT}\mathbf{Q}_{NT}^{-1}\mathbf{Q}'_{h\bar{x},N}\mathbf{Q}_{hh,N}^{-1} \\ -\mathbf{Q}_{hh,N}^{-1}\mathbf{Q}_{h\bar{x},N}\mathbf{Q}^{-1} & \mathbf{Q}_{hh,N}^{-1}\mathbf{Q}_{h\bar{x},N}\mathbf{Q}_{NT}^{-1}\mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1} & \mathbf{Q}_{hh,N}^{-1} + \mathbf{Q}_{hh,N}^{-1}\mathbf{Q}_{h\bar{x},N}\mathbf{Q}_{NT}^{-1}\mathbf{Q}'_{h\bar{x},N}\mathbf{Q}_{hh,N}^{-1} \end{pmatrix}$$

where $\mathbf{Q}_{NT}$ is given by (32). Hence

$$\tilde{\boldsymbol{\beta}} = \mathbf{Q}_{NT}^{-1}\left(\mathbf{q}_{p,NT} - \mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1}\mathbf{Q}_{z\bar{y},N} - \mathbf{Q}'_{h\bar{x},N}\mathbf{Q}_{hh,N}^{-1}\mathbf{Q}_{h\bar{y},N}\right).$$

But using the results in the lemma A.2 we note that

$$\begin{aligned}
& \mathbf{q}_{p,NT} - \mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1}\mathbf{Q}_{z\bar{y},N} - \mathbf{Q}'_{h\bar{x},N}\mathbf{Q}_{hh,N}^{-1}\mathbf{Q}_{h\bar{y},N} \\
= \; & \mathbf{Q}_{p,NT}\hat{\boldsymbol{\beta}} + \mathbf{Q}'_{z\bar{x},N}\hat{\boldsymbol{\gamma}} + \mathbf{Q}'_{h\bar{x},N} - \mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1}\left(\mathbf{Q}_{z\bar{x},N}\hat{\boldsymbol{\beta}} + \mathbf{Q}_{zz,N}\hat{\boldsymbol{\gamma}}\right) \\
& -\mathbf{Q}'_{h\bar{x},N}\mathbf{Q}_{hh,N}^{-1}\left(\mathbf{Q}_{h\bar{x},N}\hat{\boldsymbol{\beta}} + \mathbf{Q}_{hh,N}\right) \\
= \; & \mathbf{Q}_{p,NT}\hat{\boldsymbol{\beta}} + \mathbf{Q}'_{z\bar{x},N}\hat{\boldsymbol{\gamma}} + \mathbf{Q}'_{h\bar{x},N} - \mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1}\mathbf{Q}_{z\bar{x},N}\hat{\boldsymbol{\beta}} - \mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1}\mathbf{Q}_{zz,N}\hat{\boldsymbol{\gamma}} \\
& -\mathbf{Q}'_{h\bar{x},N}\mathbf{Q}_{hh,N}^{-1}\mathbf{Q}_{h\bar{x},N}\hat{\boldsymbol{\beta}} - \mathbf{Q}'_{h\bar{x},N}\mathbf{Q}_{hh,N}^{-1}\mathbf{Q}_{hh,N} \\
= \; & \mathbf{Q}_{p,NT}\hat{\boldsymbol{\beta}} + \mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1}\left(\mathbf{Q}_{z\bar{y},N} - \mathbf{Q}_{z\bar{x},N}\hat{\boldsymbol{\beta}}\right) + \mathbf{Q}'_{h\bar{x},N} - \mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1}\mathbf{Q}_{z\bar{x},N}\hat{\boldsymbol{\beta}} \\
& -\mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1}\mathbf{Q}_{zz,N}\mathbf{Q}_{zz,N}^{-1}\left(\mathbf{Q}_{z\bar{y},N} - \mathbf{Q}_{z\bar{x},N}\hat{\boldsymbol{\beta}}\right) - \mathbf{Q}'_{h\bar{x},N}\mathbf{Q}_{hh,N}^{-1}\mathbf{Q}_{h\bar{x},N}\hat{\boldsymbol{\beta}} \\
& -\mathbf{Q}'_{h\bar{x},N}\mathbf{Q}_{hh,N}^{-1}\mathbf{Q}_{hh,N} \\
= \; & \left(\mathbf{Q}_{p,NT} - \mathbf{Q}'_{z\bar{x},N}\mathbf{Q}_{zz,N}^{-1}\mathbf{Q}_{z\bar{x},N} - \mathbf{Q}'_{h\bar{x},N}\mathbf{Q}_{hh,N}^{-1}\mathbf{Q}_{h\bar{x},N}\right)\hat{\boldsymbol{\beta}} = \mathbf{Q}_{NT}\hat{\boldsymbol{\beta}}.
\end{aligned}$$

Hence, given that $\mathbf{Q}_{NT}$ is non-singular by assumption then $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$. Using this result in (A.4) and (A.5) now establishes that $\tilde{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}$, and $\tilde{\delta} = 1$, as required.

A.6 Proof of proposition 4

Denote the residuals from the OLS regression of by $\mathring{h}_i$ and note that in this case the FEVD estimators are obtained by application of the pooled OLS procedure to the following regression

$$y_{it} = \tilde{\alpha} + \mathbf{x}'_{it}\tilde{\boldsymbol{\beta}} + \mathbf{z}'_i\tilde{\boldsymbol{\gamma}} + \tilde{\delta}\mathring{h}_i + \tilde{\zeta}_{it},$$

where

$$\mathring{h}_i = \bar{\hat{u}}_i - \mathbf{z}'_i\mathring{\boldsymbol{\gamma}},$$

and $\tilde{\zeta}_{it}$ are the residuals from the pooled OLS regression. Recall also that when an intercept is included in the second step regression we have

$$\hat{h}_i = \bar{\hat{u}}_i - \hat{a}_\gamma - \mathbf{z}'_i\hat{\boldsymbol{\gamma}}.$$

Hence,

$$\mathring{h}_i + \mathbf{z}_i'\mathring{\boldsymbol{\gamma}} = \hat{h}_i + \hat{a}_\gamma + \mathbf{z}_i'\hat{\boldsymbol{\gamma}}$$

Using this result to substitute $\mathring{h}_i$ in terms of $\hat{h}_i$ we obtain

$$y_{it} = \tilde{\alpha} + \mathbf{x}_{it}'\tilde{\boldsymbol{\beta}} + \mathbf{z}_i'\tilde{\boldsymbol{\gamma}} + \tilde{\delta}\left(\hat{h}_i + \hat{a}_\gamma + \mathbf{z}_i'\hat{\boldsymbol{\gamma}} - \mathbf{z}_i'\mathring{\boldsymbol{\gamma}}\right) + \tilde{\zeta}_{it},$$

or

$$y_{it} = \left(\tilde{\alpha} + \tilde{\delta}\hat{a}_\gamma\right) + \mathbf{x}_{it}'\tilde{\boldsymbol{\beta}} + \mathbf{z}_i'\left(\tilde{\boldsymbol{\gamma}} + \hat{\boldsymbol{\gamma}} - \mathring{\boldsymbol{\gamma}}\right) + \tilde{\delta}\hat{h}_i + \tilde{\zeta}_{it}.$$

This is the same regression estimated in the third step of the FEVD procedure when an intercept term is included in the second stage, and the results of proposition 3 applies directly and we must have

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}},\ \tilde{\delta} = 1,$$

and

$$\left(\tilde{\boldsymbol{\gamma}} + \tilde{\delta}\hat{\boldsymbol{\gamma}} - \tilde{\delta}\mathring{\boldsymbol{\gamma}}\right) = \hat{\boldsymbol{\gamma}},$$

Hence, $\tilde{\boldsymbol{\gamma}} = \mathring{\boldsymbol{\gamma}}$.

To derive the bias of $\mathring{\boldsymbol{\gamma}}$ as an estimator of $\boldsymbol{\gamma}$, note that

$$\begin{aligned}
\mathring{\boldsymbol{\gamma}} &= \left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\sum_{i=1}^{N}\mathbf{z}_i\bar{\hat{u}}_i = \left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\sum_{i=1}^{N}\mathbf{z}_i\left(\bar{y}_i - \bar{\mathbf{x}}_i'\hat{\boldsymbol{\beta}}\right)\\
&= \boldsymbol{\gamma} + \left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\sum_{i=1}^{N}\mathbf{z}_i\left[\alpha + \eta_i + \bar{u}_i - \bar{\mathbf{x}}_i'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right],
\end{aligned}$$

Hence,

$$E\left(\mathring{\boldsymbol{\gamma}}\,|\mathbf{Z}\right) = \boldsymbol{\gamma} + \alpha\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\bar{\mathbf{z}},$$

and $\mathring{\boldsymbol{\gamma}}$ is an unbiased estimator of $\boldsymbol{\gamma}$, if $\alpha E\left[\left(N^{-1}\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right)^{-1}\bar{\mathbf{z}}\right] = \mathbf{0}$. Note that the bias term does not vanish even for N sufficiently large if $\alpha E(\mathbf{z}_i) \neq \mathbf{0}$, for at least one i.

A.7 Covariance matrix of the HT estimator in the case where the fixed effects are heteroskedastic and cross-sectionally correlated

Starting with (42), and using (35) we have

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{HT} &= \mathbf{Q}^{-1}\left[\mathbf{W}'\boldsymbol{\Omega}^{-1/2}\mathbf{P}_A\boldsymbol{\Omega}^{-1/2}\mathbf{y}\right]\\
&= \boldsymbol{\theta} + \mathbf{Q}^{-1}\left[\mathbf{W}'\boldsymbol{\Omega}^{-1/2}\mathbf{P}_A\boldsymbol{\Omega}^{-1/2}\mathbf{u}\right],
\end{aligned}$$

where $\mathbf{Q} = \mathbf{W}'\boldsymbol{\Omega}^{-1/2}\mathbf{P}_A\boldsymbol{\Omega}^{-1/2}\mathbf{W}$, $\mathbf{W} = [(\boldsymbol{\tau}_N \otimes \boldsymbol{\tau}_T), \mathbf{X}, (\mathbf{Z} \otimes \boldsymbol{\tau}_T)]$, and $\mathbf{u} = (\boldsymbol{\eta} \otimes \boldsymbol{\tau}_T) + \boldsymbol{\varepsilon}$. Hence, conditional on $\mathbf{W}$ we have

$$Var\left(\hat{\boldsymbol{\theta}}_{HT}\right) = \mathbf{Q}^{-1}\left[\mathbf{W}'\boldsymbol{\Omega}^{-1/2}\mathbf{P}_A\Omega^{-1/2}Var\left(\mathbf{u}\right)\boldsymbol{\Omega}^{-1/2}\mathbf{P}_A\boldsymbol{\Omega}^{-1/2}\mathbf{W}\right]\mathbf{Q}^{-1}. \tag{A.37}$$

where

$$Var\left(\mathbf{u}\right) = Var\left((\boldsymbol{\eta} \otimes \boldsymbol{\tau}_T) + \boldsymbol{\varepsilon}\right) = \mathbf{V}_\eta \otimes \boldsymbol{\tau}_T\boldsymbol{\tau}_T' + \left(\mathbf{I}_N \otimes \mathbf{I}_T\right)\sigma_\varepsilon^2,$$

and $\mathbf{V}_\eta = E(\boldsymbol{\eta}\boldsymbol{\eta}')$. Recalling that $\boldsymbol{\Omega}^{-1/2} = \frac{1}{\sigma_\varepsilon}\left[\mathbf{I}_N \otimes \mathbf{M}_T + \varphi \mathbf{I}_N \otimes (\mathbf{I}_T - \mathbf{M}_T)\right]$ with $\varphi = \sigma_\varepsilon / \sqrt{\sigma_\varepsilon^2 + T\sigma_\eta^2}$, then we have

$$\begin{aligned}
& Var\,(\mathbf{u}) = \boldsymbol{\Omega}^{-1/2} Var\,((\boldsymbol{\eta} \otimes \boldsymbol{\tau}_T) + \boldsymbol{\varepsilon})\,\boldsymbol{\Omega}^{-1/2} \\
= \; & \frac{\varphi^2}{\sigma_\varepsilon^2}\left(\mathbf{V}_\eta \otimes \boldsymbol{\tau}_T \boldsymbol{\tau}_T'\right) + \mathbf{I}_N \otimes \mathbf{M}_T + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\eta^2}\mathbf{I}_N \otimes (\mathbf{I}_T - \mathbf{M}_T) \\
= \; & \frac{T}{\sigma_\varepsilon^2 + T\sigma_\eta^2}\left(\mathbf{V}_\eta \otimes \frac{1}{T}\boldsymbol{\tau}_T \boldsymbol{\tau}_T'\right) + \mathbf{I}_N \otimes \mathbf{M}_T + \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\eta^2}\mathbf{I}_N \otimes (\mathbf{I}_T - \mathbf{M}_T) \\
= \; & \mathbf{I}_N \otimes \mathbf{I}_T - \left(1 - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\eta^2}\right)\mathbf{I}_N \otimes \frac{1}{T}\boldsymbol{\tau}_T \boldsymbol{\tau}_T' + \frac{T}{\sigma_\varepsilon^2 + T\sigma_\eta^2}\left(\mathbf{V}_\eta \otimes \frac{1}{T}\boldsymbol{\tau}_T \boldsymbol{\tau}_T'\right) \\
= \; & \mathbf{I}_N \otimes \mathbf{I}_T + \frac{T}{\sigma_\varepsilon^2 + T\sigma_\eta^2}\left((\mathbf{V}_\eta - \sigma_\eta^2 \mathbf{I}_N) \otimes \frac{1}{T}\boldsymbol{\tau}_T \boldsymbol{\tau}_T'\right).
\end{aligned}$$

Using this result in (A.37) and after some algebra we obtain (conditional on $\mathbf{W}$)

$$\begin{aligned}
Var\left(\hat{\boldsymbol{\theta}}_{HT}\right) &= \mathbf{Q}^{-1}\left[\mathbf{W}'\boldsymbol{\Omega}^{-1/2}\mathbf{P}_A\left(\mathbf{I}_N \otimes \mathbf{I}_T + \frac{T}{\sigma_\varepsilon^2 + T\sigma_\eta^2}\left((\mathbf{V}_\eta - \sigma_\eta^2 \mathbf{I}_N) \otimes \frac{1}{T}\boldsymbol{\tau}_T \boldsymbol{\tau}_T'\right)\right)\mathbf{P}_A \boldsymbol{\Omega}^{-1/2}\mathbf{W}\right]\mathbf{Q}^{-1} \\
&= \mathbf{Q}^{-1} + \left(\frac{T}{\sigma_\varepsilon^2 + T\sigma_\eta^2}\right)\mathbf{Q}^{-1}\left[\mathbf{W}'\Omega^{-1/2}\mathbf{P}_A\left((\mathbf{V}_\eta - \sigma_\eta^2 \mathbf{I}_N) \otimes \frac{1}{T}\boldsymbol{\tau}_T \boldsymbol{\tau}_T'\right)\mathbf{P}_A \boldsymbol{\Omega}^{-1/2}\mathbf{W}\right]\mathbf{Q}^{-1}.
\end{aligned}$$

References

Amemiya, T. and T. MaCurdy (1986). Instrumental-variable estimation of an error-component model. *Econometrica 54*, 869–880.

Arellano, M. (1987). Computing roboust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics 49*, 431–434.

Baltagi, B. and G. Bresson (2012). A robust hausman-taylor estimator. In *Badi H. Baltagi, R. Carter Hill, Whitney K. Newey, Halbert L. White (ed.) Essays in Honor of Jerry Hausman (Advances in Econometrics, Volume 29), pp.175-214.* Emerald Group Publishing Limited.

Beck, N. (2011). Of fixed-effects and time-invariant variables. *Political Analysis 19*, 119–122.

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

Breusch, T., M. Ward, H. Nguyen, and T. Kompas (2011a). Fevd: Just iv or just mistaken? *Political Analysis 19*, 165–169.

Breusch, T., M. Ward, H. Nguyen, and T. Kompas (2011b). On the fixed-effects vector decomposition. *Political Analysis 19*, 123–134.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press, Oxford.

Greene, W. (2011a). Fixed effects vector decomposition: A magical solution to the problem of time-invariant variables in fixed effects models? *Political Analysis 19*, 135–146.

Greene, W. (2011b). Reply to rejoinder by plumper and troeger. *Political Analysis 19*, 170–172.

Hausman, B. Y. J. A. and W. Taylor (1981). Panel data and unobservable individual effects. *Econometrica 49*, 1377–1398.

Im, K., S. Ahn, P. Schmidt, and J. Wooldridge (1999). Efficient estimation of panel data models with strictly exogenous explanatory variables. *Journal of Econometrics 93*, 177–201.

Magnus, J. R. and H. Neudecker (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, New York.

Newey, W. K. and F. Windmeijer (2009). Generalized method of moments with many weak moment conditions. *Econometrica 77*, 687–719.

Pesaran, M. (2015). *Time series and panel data econometrics*. Oxford University Press, Oxford.

Plumper, T. and V. E. Troeger (2007). Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis 15*, 124–139.

Plumper, T. and V. E. Troeger (2011). Fixed-effects vector decomposition: Properties, reliability, and instruments. *Political Analysis 19*, 147–164.

Vella, F. and M. Verbeek (1998). Whose wages do unions raise? a dynamic model of unionism and wage rate determination for yound men. *Journal of Applied Econometrics 13*, 163–183.

Table 1: Bias, RMSE, size and power of FEF and FEVD estimators for γ_1 in the case of DGP with exogenous time-invariant regressors (DGP A) and homoskedastic and serially uncorrelated errors (Case 1)

	T	3			5			10		
N		FEF	FEVD		FEF	FEVD		FEF	FEVD	
			without	with		without	with		without	with
500	Bias	0.0005	-0.1672	0.0005	-0.0001	-0.1732	-0.0001	0.0012	-0.1829	0.0012
	RMSE	0.0420	0.1830	0.0420	0.0391	0.1893	0.0391	0.0367	0.1972	0.0137
	size	5.7%	91%	44%	5.7%	95%	48%	6.2%	98%	55%
	power	22%	75%	69%	27%	83%	72%	31%	91%	82%
1000	Bias	0.0002	-0.1791	0.0002	0.0012	-0.1855	0.0012	-0.0005	-0.1950	-0.0005
	RMSE	0.0275	0.1864	0.0275	0.0257	0.1927	0.0257	0.0244	0.2016	0.0244
	size	4.7%	99%	40%	4.8%	100%	45%	5.3%	100%	53%
	power	42%	95%	86%	50%	96%	90%	52%	98%	92%
2000	bias	-0.0011	-0.1562	-0.0011	0.0001	-0.1612	0.0001	0.0001	-0.1674	0.00001
	RMSE	0.0203	0.1606	0.0203	0.0194	0.1649	0.0194	0.0181	0.1707	0.0181
	size	5.1%	100%	44%	6.3%	100%	48%	6.2%	100%	57%
	power	68%	97%	95%	76%	99%	98%	80%	100%	99%

Notes: 1. Size is calculated under $\gamma_1^{(0)} = 1$, and power under $\gamma_1^{(1)} = 0.95$.

2. The number of replication is set at $R = 1000$, and the 95% confidence interval for size 5% is [3.6%, 6.4%].

3. For FEVD estimators, "with" refers to the FEVD estimator when an intercept is included in the second step, and "without" refers to the case where the FEVD estimator is computed without an intercept.

Table 2: Bias, RMSE, size and power of FEF and FEVD estimators for γ_1 in the case of DGP with exogenous time-invariant regressors (DGP A) and heteroskedastic and serially uncorrelated errors (Case 2)

	T	3			5			10		
N		FEF	FEVD		FEF	FEVD		FEF	FEVD	
			without	with		without	with		without	with
500	Bias	0.0006	-0.1665	0.0006	0.0006	-0.1698	0.0006	0.0018	-0.1818	0.0018
	RMSE	0.0433	0.1829	0.0433	0.0375	0.1851	0.0375	0.0169	0.1946	0.0169
	size	6.5%	95%	44%	5.1%	97%	46%	41.%	99%	54%
	power	23%	85%	69%	27%	88%	74%	33%	95%	84%
1000	Bias	-0.0004	-0.1813	-0.0004	-0.0014	-0.1885	-0.0014	0.0006	-0.1975	0.0006
	RMSE	0.0282	0.1891	0.0282	0.0261	0.1954	0.0261	0.0256	0.2041	0.0256
	size	4.3%	100%	42%	4.6%	100%	47%	6%	100%	55%
	power	39%	97%	84%	47%	100%	87%	54%	100%	92%
2000	bias	0.0003	-0.1540	0.0003	0.0009	-0.1591	0.0009	-0.0012	-0.1697	-0.0012
	RMSE	0.0197	0.1584	0.0197	0.0184	0.1628	0.0184	0.0175	0.1732	0.0175
	size	4.8%	100%	42%	5.1%	100%	45%	4.8%	100%	52%
	power	71%	100%	97%	78%	100%	98%	80%	100%	99%

Notes: See notes 1-3 of Table 1.

Table 3: Bias, RMSE, size and power of FEF and FEVD estimators for γ_1 in the case of DGP with exogenous time-invariant regressors (DGP A) and serially correlated errors (Case 3)

N	T	3			5			10		
		FEF	FEVD		FEF	FEVD		FEF	FEVD	
			without	with		without	with		without	with
500	Bias	0.0016	-0.1668	0.0016	-0.0003	-0.1721	-0.0003	-0.0008	-0.1850	-0.0008
	RMSE	0.0419	0.1836	0.0419	0.0400	0.1874	0.0400	0.0371	0.1998	0.0371
	size	3.7%	91%	58%	4.2%	94%	60%	4.6%	98%	64%
	power	22%	76%	78%	23%	84%	80%	26%	92%	87%
1000	Bias	-0.0024	-0.1861	-0.0024	-0.0013	-0.1888	-0.0013	-0.0002	-0.1977	-0.0002
	RMSE	0.0299	0.1936	0.0299	0.0288	0.1961	0.0288	0.0282	0.2049	0.0282
	size	5.4%	100%	56%	5.5%	100%	58%	5.9%	100%	64%
	power	32%	96%	87%	39%	98%	89%	46%	100%	92%
2000	bias	-0.0004	-0.1548	-0.0004	0.0007	-0.1607	0.0007	0.0005	-0.1658	0.0005
	RMSE	0.0223	0.1590	0.0223	0.0210	0.1646	0.0210	0.0189	0.1694	0.0189
	size	5.6%	100%	60%	6%	100%	62%	3.4%	100%	66%
	power	62%	98%	96%	69%	100%	98%	74%	100%	100%

Notes: See notes 1-3 of Table 1.

Table 4: Bias, RMSE, size and power of FEF-IV and HTM estimators for γ_1 in the DGP with endogenous time-invariant regressors (DGP C) and homoskedastic and serially uncorrelated errors (Case 1)

N	T	3		5		10	
		FEF-IV	HTM	FEF-IV	HTM	FEF-IV	HTM
500	Bias	0.0032	0.0024	-0.0003	-0.0006	0.0016	0.0015
	RMSE	0.0490	0.0486	0.0434	0.0433	0.0419	0.0418
	size	5.9%	5.9%	4.6%	4.7%	5.4%	5.1%
	power	21%	20%	22%	20%	24%	23%
1000	Bias	0.0005	0.0000	0.0000	-0.0002	0.0013	0.0012
	RMSE	0.0335	0.0334	0.0315	0.0315	0.0288	0.0287
	size	5%	5.2%	5.3%	5%	4.8%	4.6%
	power	32%	32%	38%	36%	41%	42%
2000	Bias	-0.0001	-0.0003	0.0010	0.0009	-0.0019	-0.0020
	RMSE	0.0235	0.0235	0.0206	0.0205	0.0184	0.0184
	size	4.9%	4.4%	3.3%	3.3%	4.4%	4.6%
	power	59%	59%	66%	66%	67%	67%

Notes: 1. Size is calculated under $\gamma_1^{(0)} = 1$, and power under $\gamma_1^{(1)} = 0.95$.
2. The number of replication is set at $R = 1000$, and the 95% confidence interval for size 5% is [3.6%, 6.4%].
3. "FEF-IV" refers to the FEF-IV estimation, "HTM" refers to the modified HT estimation.

Table 5: Bias, RMSE, size and power of FEF-IV and HTM estimators for γ_1 in the DGP with endogenous time-invariant regressors (DGP C) and heteroskedastic and serially uncorrelated errors (Case 2)

	T	3		5		10	
N		FEF-IV	HTM	FEF-IV	HTM	FEF-IV	HTM
500	Bias	0.0006	-0.0002	-0.0012	-0.0015	-0.0015	-0.0016
	RMSE	0.0449	0.0446	0.0437	0.0436	0.0408	0.0407
	size	4.1%	3.6%	5.2%	4.6%	4.6%	4.2%
	power	18%	16%	21%	20%	23%	22%
1000	Bias	0.0009	0.0004	-0.0016	-0.0018	-0.0011	-0.0011
	RMSE	0.0327	0.0325	0.0313	0.0313	0.0295	0.0294
	size	4.9%	4.6%	5.2%	4.9%	4%	4.1%
	power	34%	34%	37%	37%	41%	41%
2000	Bias	-0.0001	-0.0004	0.0009	0.0008	0.0009	0.0009
	RMSE	0.0233	0.0232	0.0221	0.0221	0.0204	0.0204
	size	4.6%	4.4%	5.7%	5.4%	5%	4.6%
	power	59%	58%	64%	62%	68%	67%

Notes: See notes 1-3 of Table 4.

Table 6: Bias, RMSE, size and power of FEF-IV and HTM estimators for γ_1 in the DGP with endogenous time-invariant regressors (DGP C) and serially correlated errors (Case 3)

	T	3		5		10	
N		FEF-IV	HTM	FEF-IV	HTM	FEF-IV	HTM
500	Bias	0.0000	-0.0004	0.0001	-0.0001	-0.0001	-0.0002
	RMSE	0.0498	0.0497	0.0484	0.0483	0.0448	0.0448
	size	4.3%	4.6%	4.4%	4.7%	3.4%	3.8%
	power	17%	16%	18%	18%	21%	21%
1000	Bias	-0.0004	-0.0006	-0.0004	-0.0005	0.0002	0.0002
	RMSE	0.0362	0.0362	0.0350	0.0350	0.0317	0.0317
	size	4.5%	4.3%	6.3%	6.3%	4.5%	4.4%
	power	30%	29%	30%	30%	34%	34%
2000	Bias	0.0006	0.0005	-0.0009	-0.0009	-0.0007	-0.0007
	RMSE	0.0253	0.0252	0.0245	0.0245	0.0231	0.0231
	size	4.9%	5%	5.3%	5.1%	4.2%	4.2%
	power	51%	51%	53%	52%	55%	55%

Notes: See notes 1-3 of Table 4.