Instrumental Variable Estimation of the Correlated Random Coefficient Model with Panel Data

Rembert De Blander
CES - K.U.Leuven

February 26, 2006

Abstract

This paper considers some estimation methods for the Correlated Random Coefficient (CRC) model when panel data are available, which allows for estimation of the average treatment effect by IV as described by Wooldridge (2003), using Hausman and Taylor (1981) style instruments.

Keywords: Average treatment effect; Correlated random coefficient model; Unobserved heterogeneity

JEL classification: C21

1 Introduction

For a considerable time, different estimation procedures have been proposed for the Correlated Random Coefficient Model, albeit not always under the last name. Whenever a choice is an optimal action, instead of an exogenous datum, the subsample of individuals choosing a particular option is nonrandom. Because the non-optimal decisions cannot be observed, the impact of this selection will not be identical for everyone. More strongly, it has been demonstrated that the optimality of a choice can imply its correlation with its effect (see Garen (1984) and Card (2001) in the context of returns to schooling). Different procedures have been suggested to estimate the mean return to such a choice variable, called the average treatment effect (ATE), a term which is quite a misnomer in an economic context.

Garen (1984) assumes that the decision of interest is continuous and is solely determined by exogenous variables in a secondary equation. This author proposes an alternative estimator of the primary equation that includes both the residual of the secondary equation and the cross-product of this residual with the choice variable. This two-step procedure is termed the selectivity bias (SB) method. Heckman and Vytlacil (1998) propose a two-stage plug-in estimator using the product of the predicted values of the choice variable with the observed exogenous determinants of its return. Wooldridge (2003a) shows that a standard instrumental variables estimator where the set of instruments consists of the cross-products of the

instruments of the choice variable with the exogenous determinants of its return, consistently estimates the ATE. Wooldridge (2003b) discusses the robustness of the fixed effects (FE) and the first difference (FD) estimators for estimation of the ATE in CRC panel data models with individual-specific slopes. Belzil and Hansen (2002) conduct a structural analysis of the CRC wage regression model with individual-specific returns to schooling and experience.

On the other hand, unobservable individual- and period-specific effects can be controlled for with panel data. In increasing order of efficiency and strength of assumptions, Hausman and Taylor (1981) (HT), Amemiya and MacCurdy (1986) (AM) and Breusch, Mizon and Schmidt (1989) (BMS) proposed instrumental-variable estimation of a linear static panel data model when individual effects are correlated with a subset of the regressors. These respective estimators are compared using a returns to schooling example by Cornwell and Rupert (1988) and by Baltagi and Khanti-Akom (1990). Wyhowski (1994) presents estimators for generalized HT, AM and BMS models with two-way error components.

Random coefficient models, finally, are widely studied, particularly since the late sixties (Rao (1965)), mostly from the point of view of efficiency (Hildreth and Houck (1968)). (Swamy (1970)) and Hsiao (1975, 1974) propose estimators for the linear model with random coefficients adapted to panel data with cross-sectional, time effects or both.

In this paper a CRC model for linear static panel data is considered, where each random slope can be written as a linear function of some variables and an additive error term that consists of three components: an individual-specific, a period-specific and an idiosyncratic component. Futhermore, the correlation between the error terms of the random coefficients and the regressors can occur through any of the aforementioned components. I first propose to estimate the ATE by standard IV methods, using as instruments cross-products of the set of HT instruments of the model variables and the set of HT instruments of the slope variables, after removal of product variables that occur more than once. Furthermore, the properties of the Aitken IV and feasible Aitken IV estimators are studied.

The proposed estimators extend existing estimation methods in several ways. First of all Woolridge's (2003a) IV estimator for the CRC model in cross-sections is extended to panel data, whereby the specification of the correlation between regressors and random coefficients is kept as broad as possible. Secondly, as far as the constant term is also associated with a random and correlated coefficient, the two-way unobservable effects methodology of Wyhowski (1994) is extended in such a way that also correlation between some variables and the idiosyncratic error terms is allowed. Finally, the proposed estimator improves upon Hsiao's (1975) random coefficient estimator in that some degree of correlation between coefficients and variables is allowed and in that the coefficient are also allowed to exhibit idiosyncratic variation.

The organization of the paper is as follows. Section 2 presents the basic framework. Section 3 considers identification of the model. In section 4 estimation is considered. Finally, section 5 concludes and proofs are given in the Appendices.

2 Framework

Consider the following model

$$y_{it} = \alpha_{it} + x'_{it}\gamma_{it}, \tag{1}$$

where y_{it} is the outcome for individual i at time t, with $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where $x_{it}$is a $K \times 1$ vector of variables with associated "parameter" vector, γ_{it}, and where α_{it} is the sum of the constant term and the usual error term. The vector $(\alpha_{it}, \gamma'_{it})$ has a mean given by

$$\begin{pmatrix} \mu_\alpha \\ \mu_\gamma \end{pmatrix} = \mathrm{E}\begin{bmatrix} \alpha_{it} \\ \gamma_{it} \end{bmatrix} < \infty,$$

with μ_γ the parameters of interest. Equation (1) can consequently be written as

$$y_{it} = \mu_\alpha + x'_{it}\mu_\gamma + (\alpha_{it} - \mu_\alpha) + x'_{it}(\gamma_{it} - \mu_\gamma).$$

Suppose in addition that

$$\begin{pmatrix} \alpha_{it} - \mu_\alpha \\ \gamma_{it} - \mu_\gamma \end{pmatrix} = \begin{pmatrix} O_{1\times K} & \Psi'_\alpha \\ \Pi_\gamma & \Psi_\gamma \end{pmatrix} \begin{pmatrix} x_{it} - \mu_X \\ s_{it} - \mu_S \end{pmatrix} + \begin{pmatrix} u_{\alpha;it} \\ u_{\beta;it} \end{pmatrix}, \tag{2}$$

with

$$\begin{aligned} \mu_X &= \mathrm{E}[x_{it}], \\ \mu_S &= \mathrm{E}[s_{it}], \end{aligned}$$

where s_{it} is a $J \times 1$ vector of variables all distinct from the variables in x_{it}, Ψ_α is a $J \times 1$ parameter vector, Ψ_γ is a $K \times J$ parameter matrix, $\mathbf{O}_{P\times Q}$ is a $P \times Q$ matrix of zeroes and Π_γ is a upper triangular $K \times K$ matrix. The following assumption is made about the error term $u'_{it} = \left(u'_{\alpha;it}, u'_{\beta;it}\right)$ in the specification (2) of the random coefficients.

Assumption A. The$(K+1)\times 1$ error vector u_{it} from (2) can be written as the sum of three independent components

$$u_{it} = \lambda_i + \tau_t + \xi_{it},$$

where the components are distributed as

$$\begin{pmatrix} \lambda_i \\ \tau_t \\ \xi_{it} \end{pmatrix} \sim \mathrm{IID}\left(0; \begin{pmatrix} \Sigma_\lambda & 0 & 0 \\ 0 & \Sigma_\tau & 0 \\ 0 & 0 & \Sigma_\xi \end{pmatrix}\right).$$

Remark that λ_i, τ_t and ξ_{it} are also $(K+1)\times 1$ vectors and, consequently, Σ_λ, Σ_τ and Σ_ξ are $(K+1)\times(K+1)$ matrices. Furthermore, no assumption is made so far about the exogeneity of x_{it} and s_{it}. In the case of an error components model, Σ_λ, Σ_τ and Σ_ξ are replaced with σ^2_λ, σ^2_τ and σ^2_ξ.

Using (2)-(5), equation (1) can be rewritten as

$$\begin{aligned} y_{it} &= w'_{it}\begin{pmatrix}\mu_\alpha\\ \mu_\gamma\end{pmatrix} + w'_{it}\begin{pmatrix}O_{1\times K} & \Psi'_\alpha\\ \Pi_\gamma & \Psi_\gamma\end{pmatrix}\begin{pmatrix}x_{it}-\mu_X\\ s_{it}-\mu_S\end{pmatrix} + w'_{it}u_{it} \\ &= \mu_\alpha + x'_{it}\mu_\gamma + (s_{it}-\mu_S)'\Psi_\alpha + x'_{it}\Pi_\gamma(x_{it}-\mu_X) + x'_{it}\Psi_\gamma(s_{it}-\mu_S) + w'_{it}u_{it} \\ &= r'_{it}\varphi + w'_{it}u_{it}, \end{aligned} \tag{3}$$

with $w_{it} = (1, x'_{it})'$, $r_{it} = \left(1, x'_{it}, (s_{it}-\mu_S)', p'_{it}, q'_{it}\right)'$ and $\varphi = \left(\mu_\alpha, \mu'_\gamma, \Psi'_\alpha, \tilde{\Pi}'_\gamma, \tilde{\Psi}'_\gamma\right)'$, where $p_{it} = \text{Vech}\left[x_{it}\cdot(x_{it}-\mu_X)'\right]$, $q_{it} = \text{Vec}\left[x_{it}\cdot(s_{it}-\mu_S)'\right]$, $\tilde{\Pi}_\gamma = \text{Vech}\left[\Pi_\gamma\right]$ and $\tilde{\Psi}_\gamma = \text{Vec}\left[\Psi_\gamma\right]$[1]. Remark that the only difference between x_{it} and s_{it} is that the former vector is multiplicatively present in the error term, while the latter is not.

The matrix form of (3) can be written as

$$Y = R\varphi + W^{(D)}U^{(S)}, \tag{4}$$

where the following conventions apply: let z_{it} denote the it-th observation of a $L\times 1$vector, then we have that $Z = \left(Z'_1, \ldots, Z'_j, \ldots, Z'_N\right)'$ (a $NT\times L$ matrix), $Z_j = \left(z'_{j1}, z'_{j2}, \ldots, z'_{jT}\right)'$ (a $T\times L$ matrix), $Z^{(S)} = (z'_{11}, z'_{12}, \ldots, z'_{1T}, z'_{21}, \ldots, z'_{NT})'$ (a $LNT\times 1$ matrix), $Z^{(D)} = \text{Diag}_{i=1}^N\left[\text{Diag}_{t=1}^T\left[z'_{it}\right]\right]$, $Z^{(B)} = \text{Diag}_{i=1}^N\left[Z_i\right]$, $Z^{(T)} = \left(Z_1^{(T)\prime}, \ldots, Z_j^{(T)\prime}, \ldots, X_N^{(T)\prime}\right)'$ (all three $NT\times LNT$ matrices) and $Z_j^{(T)} = \text{Diag}_{t=1}^T\left[z'_{jt}\right]$ (a $T\times TL$ matrix), where $\text{Diag}_{j=1}^J\left[M_j\right]$ is the $(r_1+\ldots+r_J)\times(c_1+\ldots+c_J)$ block-diagonal matrix consisting of the $r_j\times c_j$ blocks M_j.

Let P_A be the projection onto the column space of the matrix A, so that $\mathcal{P}_A = A\left(A'A\right)^+A'$, where the $+$ superscript denotes the Moore-Penrose generalized inverse. Then $\mathcal{Q}_A = I - \mathcal{P}_A$ is the projection operator onto the null space of A. Letting ι_L be a $(L\times 1)$ vector of ones and $J_L = \iota_L\iota'_L$ a $(L\times L)$ matrix of ones, we can define

$$\mathcal{P}_{\iota_T} = T^{-1}J_T, \quad \mathcal{Q}_{\iota_T} = I_T - \mathcal{P}_{\iota_T}; \qquad \mathcal{P}_{\iota_N} = N^{-1}J_N, \quad \mathcal{Q}_{\iota_N} = I_N - \mathcal{P}_{\iota_N}.$$

Now define the matrix of individual dummy variables $U = I_N\otimes\iota_T$ and the matrix of period dummy variables $V = \iota_N\otimes I_T$, then we have the following four projection matrices

$$\mathcal{P}_U = T^{-1}I_N\otimes J_T = I_N\otimes\mathcal{P}_{\iota_T}, \quad \mathcal{Q}_U = I_{NT} - \mathcal{P}_U = I_N\otimes\mathcal{Q}_{\iota_T},$$

[1]The matrix operation $\text{Vec}[X]$ creates a $J^2\times 1$ column vector by stacking the columns of the $J\times J$ matrix X on top of each other. Likewise, $\text{Vech}[X]$ creates a $J(J-1)\times 1$ column vector by stacking the columns of X, starting at the diagonal elements. Finally, $\underline{\text{Vech}}[X] = \text{Vech}[X + X' - I_J\odot X]$, where $\odot$ denotes the Hadamard product.

$$\mathcal{P}_V \quad = \quad N^{-1} J_N \otimes I_T = \mathcal{P}_{\iota_N} \otimes I_T, \quad \mathcal{Q}_V = I_{NT} - \mathcal{P}_V = \mathcal{Q}_{\iota_N} \otimes I_T.$$

$\mathcal{P}_U$ transforms the data into a vector of group means, $\mathcal{Q}_U$ produces deviations from group means, $\mathcal{P}_V$ transforms the data into a vector of period means and $\mathcal{Q}_V$ produces deviations from period means. Combining both orthogonal projections, we also have that

$$X \quad = \quad \mathcal{P}_m X + \mathcal{P}_i X + \mathcal{P}_p X + \mathcal{P}_r X,$$

with $\mathcal{P}_m = \mathcal{P}_U \mathcal{P}_V$, $\mathcal{P}_i = \mathcal{P}_U \mathcal{Q}_V$, $\mathcal{P}_p = \mathcal{Q}_U \mathcal{P}_V$ and $\mathcal{P}_r = \mathcal{Q}_U \mathcal{Q}_V$, where all four terms are orthogonal to each other. These orthogonal projections allow the isolation of any possible correlation between error components and regressors in some special cases of the correlated coefficient model (Wyhowski (1994)).

3 Identification

Suppose now that we possess prior information that allows us to discern three distinct partitions of the vector x_{it}. The first partition distinguishes between the components of x_{it} along the individual-specific dimension.

Assumption B.1. X can be partitioned as

$$X \quad = \quad \left(X^a \mid X^{\dot{a}} \mid X^{\ddot{a}} \right),$$

such that

$$\begin{aligned} \mathrm{E}\left[\lambda_i \left| x_{it}^a \right.\right] &= 0, \quad \text{and} \quad \mathcal{P}_U \mathcal{Q}_V X^a \neq 0, \\ \mathrm{E}\left[\lambda_i \left| x_{it}^{\dot{a}} \right.\right] &= k_\lambda \neq 0, \quad \text{and} \quad \mathcal{P}_U \mathcal{Q}_V X^{\dot{a}} \neq 0, \\ \mathcal{P}_U \mathcal{Q}_V X^{\ddot{a}} &= 0. \end{aligned}$$

The first subset consists of k_a variables with a non-zero individual-specific component, of which λ_i is mean-independent, the second subset contains $k_{\dot{a}}$ variables with a non-zero individual-specific component, that are correlated with λ_i and in the third subset are situated $k_{\ddot{a}} = K - k_a - k_{\dot{a}}$ variables without an individual-specific component. Analogously, two more partitions are possible.

Assumption B.2. X can be partitioned as

$$X \quad = \quad \left(X^b \mid X^{\dot{b}} \mid X^{\ddot{b}} \right)$$

where

$$\begin{aligned} \mathrm{E}\left[\tau_t \left| x_{it}^{b} \right.\right] &= 0, \quad \text{and} \quad \mathcal{Q}_U \mathcal{P}_V X^{b} \neq 0, \\ \mathrm{E}\left[\tau_t \left| x_{it}^{\dot{b}} \right.\right] &= k_\tau \neq 0, \quad \text{and} \quad \mathcal{Q}_U \mathcal{P}_V X^{\dot{b}} \neq 0, \\ \mathcal{Q}_U \mathcal{P}_V X^{\ddot{b}} &= 0. \end{aligned}$$

Assumption B.3. X can be partitioned as

$$X = \left(X^{c} \mid X^{\dot{c}} \mid X^{\ddot{c}} \right)$$

with

$$\begin{aligned} \mathrm{E}\left[\xi_{it} \left| x_{it}^{c} \right.\right] &= 0, \quad \text{and} \quad \mathcal{Q}_U \mathcal{Q}_V X^{c} \neq 0, \\ \mathrm{E}\left[\xi_{it} \left| x_{it}^{\dot{c}} \right.\right] &= k_\xi \neq 0, \quad \text{and} \quad \mathcal{Q}_U \mathcal{Q}_V X^{\dot{c}} \neq 0, \\ \mathcal{Q}_U \mathcal{Q}_V X^{\ddot{c}} &= 0. \end{aligned}$$

The variables that are correlated with τ_t, respectively ξ_{it} are grouped into the subset $X^{\dot{b}}$, respectively $X^{\dot{c}}$. The subset of variables with zero time-specific, respectively idiosyncratic, component is given by $X_{it}^{\ddot{b}}$, respectively $X_{it}^{\ddot{c}}$. Considered together, these three partions define an overall partition of X with $3^3 - 1$ classes

$$X = \left(X^{abc} \mid X^{\dot{a}bc} \mid X^{\ddot{a}bc} \mid \cdots \mid X^{a\ddot{b}\ddot{c}} \mid X^{\dot{a}\ddot{b}\ddot{c}} \right).$$

Remark that $X^{\ddot{a}\ddot{b}\ddot{c}}$ is not included since it is proportional to the constant term. The subset $x_{it}^{a\dot{b}\ddot{c}}$, for instance consist of $k_{a\dot{b}\ddot{c}}$ variables with zero idiosyncratic component and non-zero individual- and time-specific components of which λ_i is mean-independent, but correlated with τ_t. In a similar way the vector s_{it} can be partitioned in 26 classes as

$$S = \left(S^{abc} \mid S^{\dot{a}bc} \mid S^{\ddot{a}bc} \mid \cdots \mid S^{a\ddot{b}\ddot{c}} \mid S^{\dot{a}\ddot{b}\ddot{c}} \right),$$

where the number of variables in S^{abc} is given by j_{abc}, the number in $S^{\dot{a}bc}$ by $j_{\dot{a}bc}$, etc. Defining the matrices A and B, with elements a_{it}, respectively b_{it}, as

$$\begin{aligned} A &= \left(\mathcal{P}_U \mathcal{Q}_V X^{a} \mid \mathcal{Q}_U \mathcal{P}_V X^{b} \mid \mathcal{Q}_U \mathcal{Q}_V X^{c} \right), \\ B &= \left(\mathcal{P}_U \mathcal{Q}_V S^{a} \mid \mathcal{Q}_U \mathcal{P}_V S^{b} \mid \mathcal{Q}_U \mathcal{Q}_V S^{c} \right), \end{aligned}$$

with

$$Z^{a} = \left(Z^{abc} \mid Z^{a\dot{b}c} \mid Z^{ab\dot{c}} \mid Z^{a\dot{b}\dot{c}} \mid Z^{a\ddot{b}c} \mid Z^{ab\ddot{c}} \mid Z^{a\ddot{b}\ddot{c}} \mid Z^{a\dot{b}\ddot{c}} \mid Z^{a\ddot{b}\dot{c}} \right),$$

$$
\begin{aligned}
Z^b &= \left(Z^{abc} \mid Z^{\grave{a}bc} \mid Z^{ab\grave{c}} \mid Z^{\grave{a}b\grave{c}} \mid Z^{\ddot{a}bc} \mid Z^{ab\ddot{c}} \mid Z^{\ddot{a}b\ddot{c}} \mid Z^{\grave{a}b\ddot{c}} \mid Z^{\ddot{a}b\grave{c}}\right), \\
Z^c &= \left(Z^{abc} \mid Z^{\grave{a}bc} \mid Z^{a\dot{b}c} \mid Z^{\grave{a}\dot{b}c} \mid Z^{\ddot{a}bc} \mid Z^{a\ddot{b}c} \mid Z^{\ddot{a}\ddot{b}c} \mid Z^{\grave{a}\ddot{b}c} \mid Z^{\ddot{a}\dot{b}c}\right),
\end{aligned}
$$

for $Z = X, S$, it holds that

$$
\mathrm{E}\left[u_{it} \left| a_{it}, b_{it}\right.\right] = 0.
$$

Reconsidering expression (3), I assume the following about the correlation between x_{it} and u_{it}.

Assumption C. *Conditional homoskedasticity of covariances* (Heckman and Vytlacil (1998), Wooldridge (2003a)), i.e.

$$
\mathrm{E}\left[x'_{it} u_{\beta;it} \left| a_{it}, b_{it}\right.\right] = \sigma_{xu}, \tag{5}
$$

where σ_{xu} is a scalar that is not a function of a_{it} nor b_{it}.

Taking into account assumption C, (3) can be rewritten as

$$
y_{it} = r'_{it}\tilde{\varphi} + \eta_{it}, \tag{6}
$$

with $\tilde{\varphi} = \left(\mu_\alpha + \sigma_{xu}, \mu'_\gamma, \Psi'_\alpha, \tilde{\Pi}'_\gamma, \tilde{\Psi}'_\gamma\right)'$ and $\eta_{it} = w'_{it} u_{it} - \sigma_{xu}$.

Define now the $k_{a^*} \equiv (1 + k_a + k_b + k_c)\left(k_a + k_b + k_c + 2\left(j_a + j_b + j_c + 1\right)\right)/2$-length column vector a^*_{it} of instruments as:

$$
a^*_{it} = \left(1, a'_{it}, b'_{it}, \mathrm{Vech}\left[a_{it} \cdot a'_{it}\right]', \mathrm{Vec}\left[a_{it} \cdot b'_{it}\right]'\right)'
$$

and make the following assumption about it.

Assumption D. It holds that $\mathrm{rank}\left[A^{*\prime} A^*\right] = k_{a^*}$.

The following proposition concerns the identification of model (6).

Proposition 1. Taking into account assumptions A-D, sufficient conditions for the identification of φ in (3), *except the constant term*, are given by

$$
\begin{aligned}
m_{\grave{a}\ddot{b}\ddot{c}} + m_{a\ddot{b}\ddot{c}} &\leq m_a, \\
m_{\ddot{a}\dot{b}\ddot{c}} + m_{\ddot{a}b\ddot{c}} &\leq m_b, \\
m_{\ddot{a}\ddot{b}\dot{c}} + m_{\ddot{a}\ddot{b}c} &\leq m_c,
\end{aligned} \tag{7}
$$

$$
2\left(m_{\ddot{a}bc} + m_{\ddot{a}\dot{b}c} + m_{\ddot{a}b\dot{c}} + m_{\ddot{a}\dot{b}\dot{c}}\right) \leq m_b + m_c,
$$

$$\begin{aligned} 2\left(m_{a\ddot{b}c}+m_{\dot{a}\ddot{b}c}+m_{a\ddot{b}\dot{c}}+m_{\dot{a}\ddot{b}\dot{c}}\right) &\leq m_a+m_c, \\ 2\left(m_{ab\ddot{c}}+m_{\dot{a}b\ddot{c}}+m_{a\dot{b}\ddot{c}}+m_{\dot{a}\dot{b}\ddot{c}}\right) &\leq m_a+m_b, \end{aligned} \tag{8}$$

and

$$\begin{aligned} & m_{\dot{a}\dot{b}\dot{c}}+\left(m_{\ddot{a}\dot{b}\dot{c}}+m_{\dot{a}\ddot{b}\dot{c}}+m_{\dot{a}\dot{b}\ddot{c}}\right)+\left(m_{\dot{a}\ddot{b}\ddot{c}}+m_{\ddot{a}\dot{b}\ddot{c}}+m_{\ddot{a}\ddot{b}\dot{c}}\right) \\ \leq\ & 2m_{abc}+\left(m_{\dot{a}bc}+m_{a\dot{b}c}+m_{ab\dot{c}}\right)+\left(m_{\ddot{a}bc}+m_{a\ddot{b}c}+m_{ab\ddot{c}}\right), \end{aligned} \tag{9}$$

for $m=j,k$.

A proof of Proposition 1 is given in the appendix. Conditions (7) state that the number of valid instruments for X, respectively S, in a particular dimension needs to be at least as high as the number of variables in X, respectively S, that have only one component in that particular direction. Conditions (8) ensure that variables with one component missing are identified, and (9) ensures that the number of valid instruments generated from X, respectively S, is larger than the number of variables in X, respectively S.

In the framework outlined above, i.e. $(\alpha_{it}-\mu_\alpha)$ is not a function of $(x_{it}-\mu_X)$ and Π_γ is a triangular matrix and under the assumption of conditional homoskedasticity of covariances, the parameter of interest, μ_γ, and all other components of φ, *except the constant term*, are identified, provided that the conditions from Proposition 1 hold. Remark that Proposition 1 insures that

$$\left(1+k_a+k_b+k_c\right)\left(1+\frac{k_a+k_b+k_c}{2}+j_a+j_b+j_c\right) \geq (1+K)\left(1+\frac{K}{2}+J\right),$$

i.e. the number of instruments is not smaller than the number of variables in model (3). Identification of the constant term is only ensured if assumption B is strengthened to $\sigma_{xu}=0$.

The model described so far in this paper encompasses some well-known models in the literature. It simplifies to a slight generalization of Hsiao's (1974) random coefficient model, when (2) simplifies to $(\alpha_{it}',\gamma_{it}')=\left(\mu_\alpha',\mu_\gamma'\right)+u_{it}'$ and when $\mathrm{E}\left[u_{it}\left|x_{it}\right.\right]=0$. If (1)-(2) is reduced to $y_{it}=\alpha_{it}$, with $\alpha_{it}-\mu_\alpha=\Psi_\alpha'\left(s_{it}-\mu_S\right)+u_{it}$, (3) collapses into a slight generalization of Wyhowski's (1994) two-way error component model.

4 Estimation

4.1 IV Estimation

A first, obvious, way to estimate (3) is simple IV estimation, using A^* as instruments, which results in

$$\hat{\varphi}_{IV} \quad = \quad \left(R'\mathcal{P}_{A^*}R\right)^{-1} R'\mathcal{P}_{A^*}Y. \tag{10}$$

The variables in R can be partitioned in the same 26 partition classes as the variables of X and S earlier on. Each partition class can be easily constructed from the classes of X and S. $R^{a\ddot{b}\ddot{c}}$, for instance, consists of $x_{it}^{a\ddot{b}\ddot{c}}$, $s_{it}^{a\ddot{b}\ddot{c}} - \mu_{S^{a\ddot{b}\ddot{c}}}$, $\text{Vech}\left[x_{it}^{a\ddot{b}\ddot{c}} \cdot \left(x_{it}^{a\ddot{b}\ddot{c}} - \mu_{X^{a\ddot{b}\ddot{c}}}\right)'\right]$ and $\text{Vec}\left[x_{it} \cdot \left(s_{it}^{a\ddot{b}\ddot{c}} - \mu_{S^{a\ddot{b}\ddot{c}}}\right)\right]$. All variables are now grouped into three sets within which the order of convergence is identical for all variables. The first group, R_i, contains at least the variables which are individual-specific, i.e. $R^{a\ddot{b}\ddot{c}}$ and $R^{\dot{a}\ddot{b}\ddot{c}}$. Similarly, R_p consists at least of the period specific variables and R_r of the idiosyncratic variables. The exact composition of each group is made explicit in Theorem 1 below. Define now $P = \text{plim}_{N,T\to\infty}\left[\left(NT\right)^{-1} R'\mathcal{P}_{A^*}R\right]$, $P_j = \text{plim}_{N,T\to\infty}\left[\left(NT\right)^{-1} R_j'\mathcal{P}_{A_j^*}R_j\right]$ for $j = i,p,r$ and $A_j^* = \mathcal{P}_j A^*$ for $j = i,p,r$. The following theorem concerns the asymptotic distribution of $\hat{\varphi}_{IV}$.

Theorem 1. Given assumptions A-D, and the sufficient identification conditions from Proposition 1, $\hat{\varphi}_{IV}$ is consistent and $\Psi\left(\hat{\varphi}_{IV} - \varphi\right) \overset{d}{\longrightarrow} N\left(0; P^{-1}\Gamma P^{-1}\right)$ as $N,T \to \infty$, where

$$\begin{aligned}
\Gamma &= \lim_{N,T\to\infty}\left[\left(NT\right)^{-2}\Psi R'\mathcal{P}_{A^*}\Omega_\eta\mathcal{P}_{A^*}R\Psi\right], \\
\Psi &= \text{Diag}_{j=1}^{3}\left[\sqrt{N_j}I_{\dim R_j}\right], \\
R &= \left(R_i, R_p, R_r\right),
\end{aligned}$$

with $R_i = \left(R^a, R^{\dot{a}}\right)$ and $R_p = \left(R^{\ddot{a}b}, R^{\ddot{a}\dot{b}}\right)$, for $N = o\left(T\right)$, with $R_i = \left(R^{a\ddot{b}}, R^{\dot{a}\ddot{b}}\right)$ and $R_p = \left(R^b, R^{\dot{b}}\right)$, if $T = o\left(N\right)$, and with $R_r = R^{\ddot{a}\ddot{b}}$.

1. If $\mathcal{P}_i H = \mathcal{P}_i W^{(D)}\left(\Lambda + \Xi\right)^{(S)}$ and $\mathcal{P}_p H = \mathcal{P}_p W^{(D)}\left(\Upsilon + \Xi\right)^{(S)}$ then $N_1 = N$, $N_2 = T$ and $N_3 = NT$. If $W^{(D)} = I_{NT}$, Γ can simplified to $\text{Diag}\left[\sigma_\lambda^2 P_i, \sigma_\tau^2 P_p, \sigma_\xi^2 P_r\right]$.

2. If $\mathcal{P}_i H = \mathcal{P}_i W^{(D)}\left(\Lambda + \Xi\right)^{(S)}$ and $\mathcal{P}_p H \neq \mathcal{P}_p W^{(D)}\left(\Upsilon + \Xi\right)^{(S)}$, then $N_1 = N_3 = N$ and $N_2 = \min\left(N,T\right)$.

3. If $\mathcal{P}_i H \neq \mathcal{P}_i W^{(D)}\left(\Lambda + \Xi\right)^{(S)}$ and $\mathcal{P}_p H = \mathcal{P}_p W^{(D)}\left(\Upsilon + \Xi\right)^{(S)}$, then $N_1 = \min\left(N,T\right)$ and $N_2 = N_3 = T$.

4. If $\mathcal{P}_i H \neq \mathcal{P}_i W^{(D)}\left(\Lambda + \Xi\right)^{(S)}$ and $\mathcal{P}_p H \neq \mathcal{P}_p W^{(D)}\left(\Upsilon + \Xi\right)^{(S)}$, then $N_1 = N_2 = N_3 = \min\left(N,T\right)$.

The condition $\mathcal{P}_p H = \mathcal{P}_p W^{(D)}\left(\Upsilon + \Xi\right)^{(S)}$ is satisfied if the only coefficients that show individual-specific variation are the ones associated with individual-specific variables or with the constant

term. In that case it holds that $\mathcal{P}_p W^{(D)} \Lambda^{(S)} = \mathcal{P}_r W^{(D)} \Lambda^{(S)} = 0$, a fact that insulates the individual-specific part of u_{it} from the period-specific and the idiosyncratic variables. Similarly, condition $\mathcal{P}_i H = \mathcal{P}_i W^{(D)} (\Lambda + \Xi)^{(S)}$ is fulfilled if coefficients exhibiting period-specific variation belong solely to period-specific variables or to the constant term. The combination of both conditions also occurs in a model where the only random coefficient is associated with the constant term, i.e. when the considered model collapses into a two-way error component model.

If Ω_η is known, the efficiency of $\hat{\varphi}_{IV}$ can be improved upon by the two-stage Aitken estimator

$$\hat{\varphi}_{2SA} = \left(R' \mathcal{P}_{A^*} \Omega_\eta^{-1} R \right)^{-1} R' \mathcal{P}_{A^*} \Omega_\eta^{-1} Y,$$

which is the best linear unbiased estimator of φ. Defining $P_\Omega = \text{plim}_{N,T\to\infty} \left[(NT)^{-1} R' \mathcal{P}_{A^*} \Omega_\eta^{-1} R \right]$ and R_i, R_p and R_r as in Theorem 1, the asymptotic distribution of $\hat{\varphi}_{2SA}$ is established in the following theorem.

Theorem 2. Under assumptions A-D, and the sufficient identification conditions from Proposition 1, $\hat{\varphi}_{2SA}$ is consistent and $\Psi_\Omega (\hat{\varphi}_{2SA} - \varphi) \overset{d}{\longrightarrow} N\left(0; P_\Omega^{-1} \Gamma_\Omega P_\Omega^{-1}\right)$ as $N, T \to \infty$, where

$$\begin{aligned} \Gamma_\Omega &= \lim_{N,T\to\infty} \left[(NT)^{-2} \Psi_\Omega R' \mathcal{P}_{A^*} \Omega_\eta^{-1} \mathcal{P}_{A^*} R \Psi_\Omega \right], \\ \Psi_\Omega &= \text{Diag}_{j=i,p,r} \left[\sqrt{N_j} I_{\dim R_j} \right] \\ R &= (R_{\Omega i}, R_{\Omega p}, R_{\Omega r}), \end{aligned}$$

with $R_{\Omega i} = \left(R^{a\ddot{b}\ddot{c}}, R^{\dot{a}\ddot{b}\ddot{c}} \right)$ and $R_{\Omega p} = \left(R^{b\ddot{c}}, R^{\dot{b}\ddot{c}} \right)$, for $N = o(T)$, with $R_{\Omega i} = (R^{a\ddot{c}}, R^{\dot{a}\ddot{c}})$ and $R_{\Omega p} = \left(R^{\ddot{a}b\ddot{c}}, R^{\ddot{a}\dot{b}\ddot{c}} \right)$, if $T = o(N)$, and with $R_{\Omega r} = (R^{\dot{c}}, R^{\dot{c}})$.

1. If $W = \iota_{NT}$, then $N_i = N$, $N_p = T$ and $N_r = NT$ and

$$P_\Omega^{-1} \Gamma_\Omega P_\Omega^{-1} = \text{Diag} \left[\sigma_\lambda^2 P_i^{-1}, \sigma_\tau^2 P_p^{-1}, \sigma_\xi^2 P_r^{-1} \right].$$

2. Otherwise, $N_i = N_p = N_r = NT$.

Case one only occurs in a model where the only random coefficient is associated with the constant term, i.e. when the considered model collapses into a two-way error component model. It is consistent with the result of Wyhowski (1994). Remark that for the error component model it holds that $\text{AVar}[\hat{\varphi}_{IV} - \varphi] \geq \text{AVar}[\hat{\varphi}_{2SA} - \varphi]$. For individual-specific, period-specific and idiosyncratic variables the consistency rates of $\hat{\varphi}_{IV}$ and $\hat{\varphi}_{2SA}$ are identical. Variables with two or more components belong to the group of variables for which the consistency rate

is *lowest* in the case of $\hat{\varphi}_{IV}$ (Theorem 1), whereas they belong to the group with the *highest* consistency rate in the case of $\hat{\varphi}_{2SA}$ (Theorem 2). Case two is the general case (Hsiao (1974)) in which we have that $\text{AVar}\left[\hat{\varphi}_{IV} - \varphi\right] > \text{AVar}\left[\hat{\varphi}_{2SA} - \varphi\right]$.

With the use of $\hat{\varphi}_{IV}$, we will now proceed to estimate $\Omega_\eta = \text{E}\left[HH'\right]$ (in the next subsection) in order to construct a feasible Aitken estimator (see subsection 4.3).

4.2 Covariance Estimation

The composite error term of (6) is given by $H = W^{(D)}U^{(S)} - \sigma_{xu}I_{NT}$. I make the following simplifying assumption regarding its covariance matrix.

Assumption E. The covariance matrix of the error term H of equation (6), obeys the following restriction

$$\text{E}\left[HH'\right] \;=\; \text{E}\left[A^{*(D)}\left(I_N \otimes J_T \otimes \tilde{\Sigma}_\lambda + J_N \otimes I_T \otimes \tilde{\Sigma}_\tau + I_{NT} \otimes \tilde{\Sigma}_\xi\right) A^{*(D)\prime}\right].$$

Defining $\hat{W} = \mathcal{P}_{A^*}W$ and $\tilde{W} = W - \hat{W}$, assumptions A-D imply that

$$\begin{aligned}\text{E}\left[HH' \mid A^*\right] \;=\;& \hat{W}^{(D)}\left(I_N \otimes J_T \otimes \Sigma_\lambda + J_N \otimes I_T \otimes \Sigma_\tau + I_{NT} \otimes \Sigma_\xi\right)\hat{W}^{(D)\prime} \\ &+\hat{W}^{(D)}\text{E}\left[UU'\tilde{W}^{(D)\prime} \mid A, B\right] + \text{E}\left[\tilde{W}^{(D)}U^{(S)}U^{(S)\prime} \mid A, B\right]\hat{W}^{(D)\prime} \\ &+\text{E}\left[\tilde{W}^{(D)}UU'\tilde{W}^{(D)\prime} \mid A, B\right] - \sigma_{xu}^2 I_{NT}. \qquad (11)\end{aligned}$$

Assumption E now imposes the additional condition that the third and fourth order conditional moments appearing in (11) reduce to polynomials of degree maximum two in the instruments. As a consequence it holds that $\text{E}\left[HH' \mid A^*\right]$ has the same structure as $\text{E}\left[HH' \mid W\right]$ would have under the assumptions of the usual (noncorrelated) random coefficient model, i.e. nonstochasticity of the regressors W (see Hsiao (1975)). The matrices of parameters that determine the variance of the disturbances are now given by the $k_{a^*} \times k_{a^*}$ matrices of $k_\eta = k_{a^*}\left(k_{a^*}+1\right)/2$ unknowns $\tilde{\Sigma}_\lambda$, $\tilde{\Sigma}_\tau$ and $\tilde{\Sigma}_\xi$, where the tilde reflects the fact that they have absorbed the influence of the higher order moments that appear in (11). I propose the folllowing procedure that only makes use of the residuals $\hat{H} = Y - R\hat{\varphi}_{IV}$ to estimate $\tilde{\Sigma}_\lambda$, $\tilde{\Sigma}_\tau$ and $\tilde{\Sigma}_\xi$ consistently.

The matrix (11) contains $NT\left(N+T-1\right)$ nonzero elements, which lead to the following set of equations. First, the NT diagonal elements are stacked as

$$H^{(\xi)} \;=\; W^{(\xi)}\left(\tilde{\Sigma}^{(\lambda)} + \tilde{\Sigma}^{(\tau)} + \tilde{\Sigma}^{(\xi)}\right) + E^{(\xi)}, \qquad (12)$$

where $H^{(\xi)} = \left(H_1^{(\xi)\prime}, H_2^{(\xi)\prime}, \ldots, H_N^{(\xi)\prime}\right)'$, $W^{(\xi)} = \left(w_{11}^{(\xi)}, w_{12}^{(\xi)}, \ldots, w_{NT}^{(\xi)}\right)'$ and $E^{(\xi)} = H^{(\xi)} - \text{E}\left[H^{(\xi)}\right]$ with $H_j^{(\xi)} = \hat{H}_j \odot \hat{H}_j = \left(\hat{h}_{j1}^2, \hat{h}_{j2}^2, \ldots, \hat{h}_{jT}^2\right)'$ $w_{it}^{(\xi)} = \underline{\text{Vech}}\left[a_{it}^* a_{it}^{*\prime}\right]$ and $\tilde{\Sigma}^{(\xi)} = \text{Vech}\left[\tilde{\Sigma}_\xi\right]$. Next, the $O_1 = NT\left(T-1\right)/2$ elements representing correlations between different observa-

tions on the same individual are stacked as

$$H^{(\lambda)} = W^{(\lambda)}\tilde{\Sigma}^{(\lambda)} + E^{(\lambda)}, \tag{13}$$

where $H^{(\lambda)} = \left(H_1^{(\lambda)}, H_2^{(\lambda)}, \ldots, H_N^{(\lambda)}\right)$, $W^{(\lambda)} = \left(W_1^{(\lambda)}, W_2^{(\lambda)}, \ldots, W_N^{(\lambda)}\right)$ and $E^{(\lambda)} = H^{(\lambda)} - \mathrm{E}\left[H^{(\lambda)}\right]$ with $H_j^{(\lambda)} = \left(\hat{h}_{j1}\hat{h}_{j2}, \hat{h}_{j1}\hat{h}_{j3}, \ldots, \hat{h}_{j,T-1}\hat{h}_{jT}\right)'$, $W_j^{(\lambda)} = \left(w_{j;12}^{(\lambda)}, w_{j;13}^{(\lambda)}, \ldots, w_{j;T-1,T}^{(\lambda)}\right)'$ $w_{j;st}^{(\lambda)} = \underline{\mathrm{Vech}}\left[a_{js}^* a_{jt}^{*\prime}\right]$ and $\tilde{\Sigma}^{(\lambda)} = \mathrm{Vech}\left[\tilde{\Sigma}_\lambda\right]$. Finally, the $O_2 = N(N-1)T/2$ elements representing correlations between different individuals observed at the same moment are stacked as

$$H^{(\tau)} = W^{(\tau)}\tilde{\Sigma}^{(\tau)} + E^{(\tau)}, \tag{14}$$

where $H^{(\tau)} = \left(H_1^{(\tau)}, H_2^{(\tau)}, \ldots, H_T^{(\tau)}\right)$, $W^{(\tau)} = \left(W_1^{(\tau)}, W_2^{(\tau)}, \ldots, W_T^{(\tau)}\right)$ and $E^{(\tau)} = H^{(\tau)} - \mathrm{E}\left[H^{(\tau)}\right]$ with $H_s^{(\tau)} = \left(\hat{h}_{1s}\hat{h}_{2s}, \hat{h}_{1s}\hat{h}_{3s}, \ldots, \hat{h}_{N-1,s}\hat{h}_{Ns}\right)'$, $W_s^{(\tau)} = \left(w_{s;12}^{(\tau)}, w_{s;13}^{(\tau)}, \ldots, w_{s;N-1,N}^{(\tau)}\right)'$ $w_{s;ij}^{(\tau)} = \underline{\mathrm{Vech}}\left[a_{is}^* a_{js}^{*\prime}\right]$ and $\tilde{\Sigma}^{(\tau)} = \mathrm{Vech}\left[\tilde{\Sigma}_\tau\right]$. All this can be summarized as

$$H^{(T)} = W^{(T)}\tilde{\Sigma}^{(T)} + E^{(T)},$$

where $H^{(T)} = \left(H^{(\lambda)\prime}, H^{(\tau)\prime}, H^{(\xi)\prime}\right)'$, $\tilde{\Sigma}^{(T)} = \left(\tilde{\Sigma}^{(\lambda)\prime}, \tilde{\Sigma}^{(\tau)\prime}, \tilde{\Sigma}^{(\xi)\prime}\right)'$, $E^{(T)} = \left(E^{(\lambda)\prime}, E^{(\tau)\prime}, E^{(\xi)\prime}\right)'$ and

$$W^{(T)} = \begin{pmatrix} W^{(\lambda)} & \mathbf{O}_{O_1\times O_3} & \mathbf{O}_{O_1\times O_3} \\ \mathbf{O}_{O_2\times O_3} & W^{(\tau)} & \mathbf{O}_{O_2\times O_3} \\ W^{(\xi)} & W^{(\xi)} & W^{(\xi)} \end{pmatrix},$$

with $O_3 = \left(K^2 + 5K + 1\right)/2$. The unknown parameters of (11) can now be estimated by

$$\hat{\tilde{\Sigma}}^{(T)} = \tilde{\Sigma}^{(T)} + \left(W^{(T)\prime}W^{(T)}\right)^{-1} W^{(T)\prime}E^{(T)}. \tag{15}$$

The proposed estimator is biased, but consistent with convergence rate given in the following proposition.

Proposition 2. The order of consistency for the proposed estimator of the variance components, $\hat{\tilde{\Sigma}}^{(T)}$, is given by

$$\begin{aligned} \sqrt{N}\left(\hat{\tilde{\Sigma}}^{(\lambda)} - \tilde{\Sigma}^{(\lambda)}\right) &= O_p(1), \\ \sqrt{T}\left(\hat{\tilde{\Sigma}}^{(\tau)} - \tilde{\Sigma}^{(\tau)}\right) &= O_p(1), \\ \sqrt{\min(N,T)}\left(\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)}\right) &= O_p(1). \end{aligned}$$

The proof is given in the appendix.

Rearranging the elements of $\hat{\tilde{\Sigma}}^{(T)}$, we obtain consistent estimates of $\tilde{\Sigma}_\lambda$, $\tilde{\Sigma}_\tau$ and $\tilde{\Sigma}_\xi$, allowing the construction of

$$\hat{\Omega}_\eta \;=\; \mathrm{E}\left[A^{*(D)}\left(I_N \otimes J_T \otimes \hat{\tilde{\Sigma}}_\lambda + J_N \otimes I_T \otimes \hat{\tilde{\Sigma}}_\tau + I_{NT} \otimes \hat{\tilde{\Sigma}}_\xi\right) A^{*(D)\prime}\right].$$

In the appendix an expression is given which simplifies the inversion of the $NT \times NT$ matrix $\hat{\Omega}_\eta$ to the inversion of the NT elements of a diagonal matrix, the inversion of a $NK \times NK$ matrix and the inversion of a $TK \times TK$ matrix, using a slight modification of an expression given by Hsiao (1974).

Note that the proposed method to estimate the covariance matrix requires only four regressions.

4.3 Feasible Aitken Estimation

Using $\hat{\Omega}_\eta$ from the previous paragraph, or any other consistent estimator of φ, the feasible two-stage Aitken estimator is given by

$$\hat{\varphi}_{F2SA} \;=\; \left(R'\mathcal{P}_{A^*}\hat{\Omega}_\eta^{-1}R\right)^{-1} R'\mathcal{P}_{A^*}\hat{\Omega}_\eta^{-1}Y,$$

for which the asymptotic distribution is given in the following theorem.

Theorem 3. Under assumptions A-E, under the sufficient identification conditions from Proposition 1 and using the covariance estimator from Proposition 2, $\Psi_{\hat{\Omega}}\left(\hat{\varphi}_{F2SA} - \varphi\right) = O_p\left(1\right)$ as $N, T \to \infty$, where

$$\begin{aligned} \Psi_{\hat{\Omega}} &= \mathrm{Diag}_{j=i,p,r}\left[\sqrt{N_j} I_{\dim R_j}\right] \\ R &= \left(R_{\hat{\Omega}i}, R_{\hat{\Omega}p}, R_{\hat{\Omega}r}\right), \end{aligned}$$

with $R_{\hat{\Omega}i} = \left(R^{a\ddot{b}\ddot{c}}, R^{\dot{a}\ddot{b}\ddot{c}}\right)$ and $R_{\hat{\Omega}p} = \left(R^{b\ddot{c}}, R^{\dot{b}\ddot{c}}\right)$, for $N = o\left(T\right)$, with $R_{\hat{\Omega}i} = \left(R^{a\ddot{c}}, R^{\dot{a}\ddot{c}}\right)$ and $R_{\hat{\Omega}p} = \left(R^{\ddot{a}b\ddot{c}}, R^{\ddot{a}\dot{b}\ddot{c}}\right)$, if $T = o\left(N\right)$, and with $R_{\hat{\Omega}r} = \left(R^{c}, R^{\dot{c}}\right)$. Furthermore, two cases are discernable.

1. If $W = \iota_{NT}$, then $N_i = N$, $N_p = T$ and $N_r = NT$ and

$$\Psi_{\hat{\Omega}}\left(\hat{\varphi}_{F2SA} - \varphi\right) \overset{d}{\longrightarrow} N\left(0; P_\Omega^{-1}\Gamma_\Omega P_\Omega^{-1}\right),$$

with $P_\Omega^{-1}\Gamma_\Omega P_\Omega^{-1} = \mathrm{Diag}\left[\sigma_\lambda^2 P_i^{-1}, \sigma_\tau^2 P_p^{-1}, \sigma_\xi^2 P_r^{-1}\right]$.

2. Otherwise, $N_i = N_p = N_r = \min\left(N^2, T^2\right)$.

The proof of Theorem 3 is given in the appendix. Part one pertains to the error component model and it states that the asymptotic distribution of the feasible estimator is identical to that of the estimator with known covariance matrix. Part two establishes the consistency rate for the feasible IV Aitken estimator that uses the covariance estimator from subsection 4.2.

The importance of Theorem 3 lies first of all, in the fact that it provides convergence rates for the feasible two-stage Aitken estimator without imposing a restriction on the ratio N/T. A second important finding is that the feasible Aitken estimator does not converge to the Aitken estimator when we consider a (correlated) random coefficient model where the random parts of the coefficients follow a two-way error specification.

In accordance with the results of Swamy (1970), Kelejian and Stephan (1983) noted that the convergence rate reported by Hsiao is counter-intuitive. As Mandy and Martins-Filho (1994) note, the asymptotic equivalence between Aitken and feasible Aitken estimators is often proved by relying on Theorem 3 of Fuller and Battese (1973), which the former authors disprove by means of a counter-example. More specifically, the conditions under which Fuller and Battese state their theorem to be valid are much to general. The asymptotic equivalence between Aitken and feasible Aitken estimators in the case of a diagonal disturbance covariance matrix (see Carroll and Ruppert (1982) and Crocket (1985)) cannot be extended unconditionally to the case of a non-diagonal covariance matrix.

Sufficient conditions for the asymptotic equivalence of $\hat{\varphi}_{2SA}$ and $\hat{\varphi}_{F2SA}$ are

$$\text{plim}_{N,T\to\infty}\left[(NT)^{-1} R'\mathcal{P}_{A^*}\left(\hat{\Omega}_\eta^{-1}-\Omega_\eta^{-1}\right)R\right] = 0 \tag{16}$$

$$\text{plim}_{N,T\to\infty}\left[(NT)^{-1} \Psi_\Omega R'\mathcal{P}_{A^*}\left(\hat{\Omega}_\eta^{-1}-\Omega_\eta^{-1}\right)H\right] = 0 \tag{17}$$

(White (1984)). Mandy and Martins-Filho (1994) list conditions under which $\text{plim}_{N,T\to\infty}\left[\hat{\tilde{\Sigma}}^{(T)}\right] = \tilde{\Sigma}^{(T)}$ (consistency of the covariance estimator) implies (16) and (17) and thus asymptotic equivalence between $\hat{\varphi}_{2SA}$ and $\hat{\varphi}_{F2SA}$. Unfortunately, one of their conditions requires the number of nonzero elements in each column (and thus row) of $\hat{\Omega}_\eta^{-1}$ to be uniformly bounded as $N,T\to\infty$. Consequently, their Theorem 1 cannot be used to prove asymptotic equivalence of $\hat{\varphi}_{2SA}$ and $\hat{\varphi}_{F2SA}$ here. For the special case of the error components model (part 1 in Theorem 3), (16) and (17) are easily proved. For the general CRC model, (16) also holds, but (17) is violated and the correct order of convergence is shown to be $\min(N^2,T^2)$.

5 Concluding Remarks

In this paper some estimators are presented for the linear CRC model with panel data, where each random slope can be written as a linear function of some variables and an additive error term that consists of three components, which can all three be correlated with some of the regressors. I propose to estimate the ATE by standard IV methods, using as instruments cross-

products of the set of HT instruments of the model variables and the set of HT instruments of the slope variables, after removal of product variables that occur more than once. Sufficient identification conditions for this estimator are obtained and it is shown to be asymptotically normal. Also the properties of the Aitken and feasible Aitken IV estimators were investigated. In the special case of the error-component model they were shown to converge to the same asymptotic distribution. In the case of a general CRC model, however, they were shown to be asymptotic nonequivalent and the rate of convergence of the feasible Aitken IV estimator was obtained.

Acknowledgments

Helpful comments by Geert Dhaene are gratefully acknowledged. All residual errors, however, are exclusively the author's responsibility. This research was supported by FWO grant # G.0198.97 and IUAP grant # P4/01 (A1281).

Appendix

A Identification

Proof of Proposition 1. A necessary condition for identifiability of γ is that $P_{A^*}R$ is of full column rank. Variables in X and $S-\mu_S$ that have only a nonzero component in one of the three (individual-specific, time-specific and idiosyncratic) dimensions need to be identified. For the individual-specific dimension a necessary condition for this is that

$$\begin{aligned} m_{a\ddot{b}\ddot{c}} + m_{\dot{a}\ddot{b}\ddot{c}} &\leq m_a \\ &= m_{abc} + m_{a\dot{b}c} + m_{ab\dot{c}} + m_{a\dot{b}\dot{c}} + m_{a\ddot{b}c} + m_{ab\ddot{c}} + m_{a\ddot{b}\ddot{c}} + m_{a\dot{b}\ddot{c}} + m_{a\ddot{b}\dot{c}}, \end{aligned}$$

for $m = j, k$. The same condition holds for such variables along other dimensions

$$\begin{aligned} m_{\ddot{a}\dot{b}\ddot{c}} &\leq m_b - m_{\ddot{a}b\ddot{c}}, \\ m_{\ddot{a}\ddot{b}\dot{c}} &\leq m_c - m_{\ddot{a}\ddot{b}c}, \end{aligned}$$

for $m = j, k$. Variables in X and $S-\mu_S$ that have one dimension missing need to be identified in one of the two other dimensions. For the variables that miss the individual-specific dimension a necessary condition for this is given by

$$2\left(m_{\ddot{a}bc} + m_{\ddot{a}\dot{b}c} + m_{\ddot{a}b\dot{c}} + m_{\ddot{a}\dot{b}\dot{c}}\right) \leq m_b + m_c,$$

for $m = j, k$ and analogously for variables missing other dimensions

$$2\left(m_{a\ddot{b}c} + m_{\dot{a}\ddot{b}c} + m_{a\ddot{b}\dot{c}} + m_{\dot{a}\ddot{b}\dot{c}}\right) \leq m_a + m_c,$$

$$2\left(m_{ab\ddot{c}}+m_{\dot{a}b\ddot{c}}+m_{a\dot{b}\ddot{c}}+m_{\dot{a}\dot{b}\ddot{c}}\right) \leq m_a+m_b,$$

for $m=j,k$. Finally, the number of variables in X and $S-\mu_S$ needs to be smaller than the number of instruments for X, respectively $S-\mu_S$,

$$\begin{aligned} K &\leq k_a+k_b+k_c, \\ J &\leq j_a+j_b+j_c, \end{aligned}$$

which, after simple algebra, leads to

$$\begin{aligned} & m_{\dot{a}\dot{b}\dot{c}}+\left(m_{\ddot{a}\dot{b}\dot{c}}+m_{\dot{a}\ddot{b}\dot{c}}+m_{\dot{a}\dot{b}\ddot{c}}\right)+\left(m_{\ddot{a}\ddot{b}\ddot{c}}+m_{\ddot{a}\dot{b}\ddot{c}}+m_{\ddot{a}\ddot{b}\dot{c}}\right) \\ \leq\ & 2m_{abc}+\left(m_{\dot{a}bc}+m_{a\dot{b}c}+m_{ab\dot{c}}\right)+\left(m_{\ddot{a}bc}+m_{a\ddot{b}c}+m_{ab\ddot{c}}\right), \end{aligned}$$

for $m=j,k$.

B Estimation

Throughout the rest of this appendix denote $\mathcal{P}_{A^*}R$ by $\hat{R}$.

B.1 IV Estimation

Proof of Theorem 1. The consistency of $\hat{\varphi}_{IV}$ follows immediately from assumption C. In order to derive the asymptotic distribution of $\hat{\varphi}_{IV}$, define

$$C = \begin{pmatrix} (NT)^{-1}\sqrt{N_1}\iota_{NT}'\mathcal{P}_iH \\ (NT)^{-1}\sqrt{N_2}\iota_{NT}'\mathcal{P}_pH \\ (NT)^{-1}\sqrt{N_3}\iota_{NT}'\mathcal{P}_rH \end{pmatrix},$$

and $\Sigma_C=\mathrm{Var}\left[CC'\right]$. From the construction of $\mathcal{P}_i$, $\mathcal{P}_p$ and $\mathcal{P}_r$ we have, in general, that

$$\mathcal{P}_iH=\mathcal{P}_iW^{(D)}\left(\Lambda+\Xi\right)^{(S)} \Leftrightarrow \mathcal{P}_rH=\mathcal{P}_rW^{(D)}\left(\Lambda+\Xi\right)^{(S)} \tag{18}$$

$$\mathcal{P}_pH=\mathcal{P}_pW^{(D)}\left(\Upsilon+\Xi\right)^{(S)} \Leftrightarrow \mathcal{P}_rH=\mathcal{P}_rW^{(D)}\left(\Upsilon+\Xi\right)^{(S)}. \tag{19}$$

The condition $\mathcal{P}_iH=\mathcal{P}_iW^{(D)}\left(\Lambda+\Xi\right)^{(S)}$ is satisfied if the only coefficients that show period-specific variation are the ones associated with period-specific variables or with the constant term. In that case $W^{(D)}\Upsilon^{(S)}$ is also period-specific and thus orthogonal to $\mathcal{P}_i$ and $\mathcal{P}_r$. Similarly, $\mathcal{P}_pH=\mathcal{P}_pW^{(D)}\left(\Upsilon+\Xi\right)^{(S)}$ is satisfied if the only coefficients that show individual-specific variation are the ones associated with individual-specific variables or with the constant term. Fulfillment of both conditions implies that only the constant term has a random coefficient associated with it, i.e. the model is an error components model.

It also holds that

$$\begin{aligned}
\lim_{N,T\to\infty} \mathrm{E}\left[\iota_{NT}' W^{(D)} \Lambda^{(S)} \Lambda^{(S)\prime} W^{(D)\prime} \iota_{NT}\right] &= O\left(NT^2\right) \\
\lim_{N,T\to\infty} \mathrm{E}\left[\iota_{NT}' W^{(D)} \Upsilon^{(S)} \Upsilon^{(S)\prime} W^{(D)\prime} \iota_{NT}\right] &= O\left(N^2T\right) \\
\lim_{N,T\to\infty} \mathrm{E}\left[\iota_{NT}' W^{(D)} \Xi^{(S)} \Xi^{(S)\prime} W^{(D)\prime} \iota_{NT}\right] &= O\left(NT\right), \qquad (20)
\end{aligned}$$

by assumption A.

In the first case, relations (18)-(19) imply that $\mathcal{P}_i H = \mathcal{P}_i W^{(D)} \left(\Lambda + \Xi\right)^{(S)}$, $\mathcal{P}_p H = \mathcal{P}_p W^{(D)} \left(\Upsilon + \Xi\right)^{(S)}$ and $\mathcal{P}_r H = \mathcal{P}_r W^{(D)} \Xi^{(S)}$ and it holds that

$$\lim_{N,T\to\infty} \left[\Sigma_C\right] = \Psi_C \begin{pmatrix} O\left(N^{-1}\right) & O\left((NT)^{-1}\right) & O\left((NT)^{-1}\right) \\ O\left((NT)^{-1}\right) & O\left(T^{-1}\right) & O\left((NT)^{-1}\right) \\ O\left((NT)^{-1}\right) & O\left((NT)^{-1}\right) & O\left((NT)^{-1}\right) \end{pmatrix} \Psi_C,$$

where $\Psi_C = \mathrm{Diag}_{j=1}^{3}\left[\sqrt{N_j}\right]$. Now, since $\hat{R} = O_p\left(1\right)$, entries in Γ corresponding to individual-specific variables will be of the same order as the $(1,1)$-th element of $\lim_{N,T\to\infty}\left[\Sigma_C\right]$, period-specific variables correspond to the $(2,2)$-th element and idiosyncratic variables to the $(3,3)$-th element. In order that $\Gamma = O\left(1\right)$, we need to impose that $N_1 = N$, $N_2 = T$ and $N_3 = NT$. Variables with more than one component belong to the group of variables for which the variance has the highest order. For instance, a variable with both an individual-specific and a period-specific component will be part of R_i if $N = o\left(T\right)$, but it will be an element of R_p if $T = o\left(N\right)$. Consequently, we have that $R_i = \left(R^{a}, R^{\dot{a}}\right)$, $R_p = \left(R^{\ddot{a}b}, R^{\ddot{a}\dot{b}}\right)$, for $N = o\left(T\right)$ and $R_i = \left(R^{\ddot{b}a}, R^{\ddot{b}\dot{a}}\right)$, $R_p = \left(R^{b}, R^{\dot{b}}\right)$, if $T = o\left(N\right)$ and always that $R_r = R^{\ddot{a}\ddot{b}}$.

The expression for Γ in the first case can be simplified, by noticing that $R'\mathcal{P}_{A^*}H = R'\mathcal{P}_{A_i^*}\left(\Lambda + \Xi\right) + R'\mathcal{P}_{A_p^*}\left(\Upsilon + \Xi\right) + R'\mathcal{P}_{A_r^*}\Xi$. Consequently, $\Gamma = \lim_{N,T\to\infty} \mathrm{Var}\left[\left(NT\right)^{-1}\Psi R'\mathcal{P}_{A^*}H\right]$ can be written as

$$\begin{aligned}
\Gamma &= \lim_{N,T\to\infty}\left[\left(NT\right)^{-2}\Psi R'\mathcal{P}_{A^*}\Omega_\eta\mathcal{P}_{A^*}R\Psi\right] \\
&= \lim_{N,T\to\infty}\left[\left(T\sigma_\lambda^2 + \sigma_\xi^2\right)\left(NT\right)^{-2}\Psi R'\mathcal{P}_{A_i^*}\mathcal{P}_i\mathcal{P}_{A_i^*}R\Psi\right] \\
&\quad + \lim_{N,T\to\infty}\left[\left(N\sigma_\tau^2 + \sigma_\xi^2\right)\left(NT\right)^{-2}\Psi R'\mathcal{P}_{A_p^*}\mathcal{P}_p\mathcal{P}_{A_p^*}R\Psi\right] \\
&\quad + \lim_{N,T\to\infty}\left[\sigma_\xi^2\left(NT\right)^{-2}\Psi R'\mathcal{P}_{A_r^*}\mathcal{P}_r\mathcal{P}_{A_r^*}R\Psi\right]
\end{aligned}$$

and further simplification results in $\Gamma = \mathrm{Diag}_{j=1}^{3}\left[M_j\right]$, with $M_1 = \sigma_\lambda^2 P_i$, $M_2 = \sigma_\tau^2 P_p$, $M_3 = \sigma_\xi^2 P_r$ where $P_j = \mathrm{plim}_{N,T\to\infty}\left[\left(NT\right)^{-1} R_j'\mathcal{P}_{A_j^*}R_j\right]$ for $j = i, p, r$.

By the same reasoning, it holds in the second case that

$$\lim_{N,T\to\infty}\left[\Sigma_C\right] = \Psi_C \begin{pmatrix} O\left(N^{-1}\right) & O\left(N^{-1}\right) & O\left(N^{-1}\right) \\ O\left(N^{-1}\right) & O\left(N^{-1}+T^{-1}\right) & O\left(N^{-1}\right) \\ O\left(N^{-1}\right) & O\left(N^{-1}\right) & O\left(N^{-1}\right) \end{pmatrix} \Psi_C,$$

$N_1 = N_3 = N$, $N_2 = \min(N,T)$ and $R_p = \left(R^b, R^{\dot{b}}\right)$, if $T = o(N)$.

In the third case we have that

$$\lim_{N,T\to\infty}\left[\Sigma_C\right] = \Psi_C \begin{pmatrix} O\left(N^{-1}+T^{-1}\right) & O\left(T^{-1}\right) & O\left(T^{-1}\right) \\ O\left(T^{-1}\right) & O\left(T^{-1}\right) & O\left(T^{-1}\right) \\ O\left(T^{-1}\right) & O\left(T^{-1}\right) & O\left(T^{-1}\right) \end{pmatrix} \Psi_C,$$

with $N_1 = \min(N,T)$, $N_2 = N_3 = N$ and $R_i = \left(R^a, R^{\dot{a}}\right)$, if $N = o(T)$.

In the fourth case, finally, it holds that $\lim_{N,T\to\infty}\left[\Sigma_C\right] = O\left(N^{-1}+T^{-1}\right)\Psi_C J_3 \Psi_C$ and thus $N_j = \min(N,T)$ for $j = 1,2,3$. Asymptotic normality follows by a slight adaptation of Theorem 4.1. in Hsiao (1974).

Proof of Theorem 2. Consistency of $\hat{\varphi}_{2SA}$ follows from the consistency of $\hat{\varphi}_{IV}$. Asymptotic normality again follows by a slight adaptation of Theorem 4.1. in Hsiao (1974). The simplification of the expression for the asymptotic variance in the case of the error component model can be seen by noticing that

$$\begin{aligned} \left(NT\right)^{-1} R'\mathcal{P}_{A^*}\Omega_\eta^{-1}R &= \left(T\sigma_\lambda^2+\sigma_\xi^2\right)^{-1}\left(NT\right)^{-1} R'\mathcal{P}_{A_i^*}R \\ &\quad + \left(N\sigma_\tau^2+\sigma_\xi^2\right)^{-1}\left(NT\right)^{-1} R'\mathcal{P}_{A_p^*}R \\ &\quad + \sigma_\xi^{-2}\left(NT\right)^{-1} R'\mathcal{P}_{A_r^*}R \end{aligned} \tag{21}$$

and

$$\begin{aligned} \left(NT\right)^{-2}\Psi_\Omega R'\mathcal{P}_{A^*}\Omega_\eta^{-1}\mathcal{P}_{A^*}R\Psi_\Omega &= \left(T\sigma_\lambda^2+\sigma_\xi^2\right)^{-1}\left(NT\right)^{-2}\Psi_\Omega R'\mathcal{P}_{A_i^*}R\Psi_\Omega \\ &\quad + \left(N\sigma_\tau^2+\sigma_\xi^2\right)^{-1}\left(NT\right)^{-2}\Psi_\Omega R'\mathcal{P}_{A_p^*}R\Psi_\Omega \\ &\quad + \sigma_\xi^{-2}\left(NT\right)^{-2}\Psi_\Omega R'\mathcal{P}_{A_r^*}R\Psi_\Omega. \end{aligned} \tag{22}$$

For individual-specific variables we have that

$$\begin{aligned} &\operatorname{plim}_{N,T\to\infty}\left[\left(NT\right)^{-1} R^{\ddot{b}\ddot{c}\prime}\mathcal{P}_{A^*}\Omega_\eta^{-1}R^{\ddot{b}\ddot{c}}\right] \\ = &\lim_{N,T\to\infty}\left[T^{-1}\sigma_\lambda^{-2}\right]\operatorname{plim}_{N,T\to\infty}\left[\left(NT\right)^{-1} R^{\ddot{b}\ddot{c}\prime}\mathcal{P}_{A^*}R^{\ddot{b}\ddot{c}}\right] \end{aligned}$$

$$\lim_{N,T\to\infty}\left[\left(NT\right)^{-2}\Psi_\Omega R^{\ddot{b}\ddot{c}\prime}\mathcal{P}_{A^*}\Omega_\eta^{-1}\mathcal{P}_{A^*}R^{\ddot{b}\ddot{c}}\Psi_\Omega\right]$$

$$= \lim_{N,T\to\infty}\left[N^{-1}T^{-2}N_1\sigma_\lambda^{-2}\right]\text{plim}_{N,T\to\infty}\left[(NT)^{-1}R^{\ddot{b}\ddot{c}\prime}\mathcal{P}_{A^*}R^{\ddot{b}\ddot{c}}\right]$$

and $N_i = N$. Similarly, $N_p = T$ and $N_r = NT$. It can also be seen from (21) and (22) that variables containing more than one component have as consistency rate the highest consistency rate of each of the components.

B.2 Covariance Estimation

Proof of Proposition 2. By construction of $W^{(T)}$ and $E^{(T)}$, we have that

$$\text{plim}_{N,T\to\infty}\left[W^{(T)\prime}W^{(T)}\right] = \begin{pmatrix} O\left(NT^2\right) & O\left(NT\right) & O\left(NT\right) \\ O\left(NT\right) & O\left(N^2T\right) & O\left(NT\right) \\ O\left(NT\right) & O\left(NT\right) & O\left(NT\right) \end{pmatrix},$$

$$\text{plim}_{N,T\to\infty}\left[\left(W^{(T)\prime}W^{(T)}\right)^{-1}\right] = \begin{pmatrix} O\left(N^{-1}T^{-2}\right) & O\left(N^{-2}T^{-2}\right) & O\left(N^{-1}T^{-2}\right) \\ O\left(N^{-2}T^{-2}\right) & O\left(N^{-2}T^{-1}\right) & O\left(N^{-2}T^{-1}\right) \\ O\left(N^{-1}T^{-2}\right) & O\left(N^{-2}T^{-1}\right) & O\left(N^{-1}T^{-1}\right) \end{pmatrix}, \quad (23)$$

and

$$\text{Var}\left[W^{(T)\prime}E^{(T)}\right] = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{12} & M_{22} & M_{23} \\ M_{13} & M_{23} & M_{33} \end{pmatrix}.$$

Now, from the definition $H_j^{(\lambda)} = \left(\hat{h}_{j1}\hat{h}_{j2}, \hat{h}_{j1}\hat{h}_{j3}, \ldots, \hat{h}_{j,T-1}\hat{h}_{jT}\right)'$ as a $T\left(T-1\right)/2\times 1$-matrix, we have that $\text{E}\left[H_i^{(\lambda)}H_j^{(\lambda)\prime}\right] - \text{E}\left[H_i^{(\lambda)}\right]\text{E}\left[H_j^{(\lambda)\prime}\right]$ has a st,uv-th element equal to

$$\begin{aligned} c_{ij;st,uv} &= \left(w_{is}'\otimes w_{ju}'\right)\text{E}\left[\left(\lambda_i\lambda_i' + \tau_s\tau_t' + \xi_{is}\xi_{it}'\right)\otimes\left(\lambda_j\lambda_j' + \tau_u\tau_v' + \xi_{ju}\xi_{jv}'\right)\right]\left(w_{it}\otimes w_{jv}\right) \\ &\quad -w_{is}'\left\{\Sigma_\lambda + \delta_{s=t}\left(\Sigma_\tau + \Sigma_\xi\right)\right\}w_{it}\cdot w_{ju}'\left\{\Sigma_\lambda + \delta_{u=v}\left(\Sigma_\tau + \Sigma_\xi\right)\right\}w_{jv}, \end{aligned}$$

from which we deduce that $\text{E}\left[H_i^{(\lambda)}H_j^{(\lambda)\prime}\right] - \text{E}\left[H_i^{(\lambda)}\right]\text{E}\left[H_j^{(\lambda)\prime}\right] = \text{Diag}_{s=1}^{(T-1)}\left[\text{Diag}_{t=s+1}^{T}\left[c_{ij;st,st}\right]\right]$, if $i\neq j$, and it is equal to a $T\left(T-1\right)/2\times T\left(T-1\right)/2$ matrix with nonzero elements if $i=j$. Thus it holds that $\text{Var}\left[W^{(\lambda)\prime}E^{(\lambda)}\right] = O\left(N\left(N-1\right)T\left(T-1\right)/2\right) + O\left(NT^2\left(T-1\right)^2/4\right)$. Since the problem at hand is symmetric with respect to N and T, we also have that

$$\text{Var}\left[W^{(\tau)\prime}E^{(\tau)}\right] = O\left(N\left(N-1\right)T\left(T-1\right)/2\right) + O\left(N^2\left(N-1\right)^2T/4\right).$$

Similarly, the s,t-th element of $\text{E}\left[H_i^{(\xi)}H_j^{(\xi)\prime}\right] - \text{E}\left[H_i^{(\xi)}\right]\text{E}\left[H_j^{(\xi)\prime}\right]$ equals

$$d_{ij;st} = \left(w_{is}'\otimes w_{jt}'\right)\text{E}\left[\left(\lambda_i\lambda_i' + \tau_s\tau_s' + \xi_{is}\xi_{is}'\right)\otimes\left(\lambda_j\lambda_j' + \tau_t\tau_t' + \xi_{jt}\xi_{jt}'\right)\right]\left(w_{is}\otimes w_{jt}\right)$$

$$-w'_{is}\left\{\Sigma_\lambda+\Sigma_\tau+\Sigma_\xi\right\}w_{is}\cdot w'_{jt}\left\{\Sigma_\lambda+\Sigma_\tau+\Sigma_\xi\right\}w_{jt},$$

which causes $\mathrm{E}\left[H^{(\xi)}H^{(\xi)\prime}\right]-\mathrm{E}\left[H^{(\xi)}\right]\mathrm{E}\left[H^{(\xi)\prime}\right]$ to have the same structure as $\mathrm{E}\left[HH'\mid A^*\right]$, resulting in $\mathrm{Var}\left[W^{(\xi)\prime}E^{(\xi)}\right]=O\left(NT\left(N+T-1\right)\right)$.

Furthermore, $\mathrm{E}\left[H_i^{(\lambda)}H_j^{(\xi)\prime}\right]-\mathrm{E}\left[H_i^{(\lambda)}\right]\mathrm{E}\left[H_j^{(\xi)\prime}\right]$, with st,u-th element

$$\begin{aligned}e_{ij;st,u} &= \left(w'_{is}\otimes w'_{ju}\right)\mathrm{E}\left[\left(\lambda_i\lambda'_i+\tau_s\tau'_t+\xi_{is}\xi'_{it}\right)\otimes\left(\lambda_j\lambda'_j+\tau_u\tau'_u+\xi_{ju}\xi'_{ju}\right)\right]\left(w_{it}\otimes w_{ju}\right)\\ &\quad -w'_{is}\left\{\Sigma_\lambda\right\}w_{it}\cdot w'_{ju}\left\{\Sigma_\lambda+\Sigma_\tau+\Sigma_\xi\right\}w_{ju},\end{aligned}$$

is equal to $\mathbf{O}_{T(T-1)\times T}$, unless $i=j$, in which case all of its elements are nonzero. Consequently we have that $\mathrm{Cov}\left[W^{(\lambda)\prime}E^{(\lambda)},W^{(\xi)\prime}E^{(\xi)}\right]=O\left(NT^2\left(T-1\right)\right)$, and, by the symmetry between N and T, $\mathrm{Cov}\left[W^{(\tau)\prime}E^{(\tau)},W^{(\xi)\prime}E^{(\xi)}\right]=O\left(N^2\left(N-1\right)T\right)$. Finally, $\mathrm{E}\left[H_i^{(\lambda)}H_t^{(\tau)\prime}\right]-\mathrm{E}\left[H_i^{(\lambda)}\right]\mathrm{E}\left[H_t^{(\tau)\prime}\right]$ has uv,kl-th element equal to

$$\begin{aligned}f_{it;uv,kl} &= \left(w'_{iu}\otimes w'_{kt}\right)\mathrm{E}\left[\left(\lambda_i\lambda'_i+\tau_u\tau'_v+\xi_{iu}\xi'_{iv}\right)\otimes\left(\lambda_k\lambda'_l+\tau_t\tau'_t+\xi_{kt}\xi'_{lt}\right)\right]\left(w_{iv}\otimes w_{lt}\right)\\ &\quad -w'_{iu}\left\{\Sigma_\lambda\right\}w_{iv}\cdot w'_{kt}\left\{\Sigma_\tau\right\}w_{lt},\end{aligned}$$

which is equal to zero, unless $(h=i\vee l=i)\wedge(u=t\vee v=t)$, i.e. at $(2N-3)(2T-3)$ places. It follows that $\mathrm{Cov}\left[W^{(\lambda)\prime}E^{(\lambda)},W^{(\tau)\prime}E^{(\tau)}\right]=O\left(N\left(2N-3\right)T\left(2T-3\right)\right)$.

From the previous we have that

$$\begin{aligned}&\mathrm{Var}\left[W^{(T)\prime}E^{(T)}\right]\\ &= \begin{pmatrix} O\left(NT^4\right)+O\left(N^2T^2\right) & O\left(N^3T\right)+O\left(NT^3\right) & O\left(NT^3\right)+O\left(N^2T\right)\\ O\left(N^3T\right)+O\left(NT^3\right) & \left(N^4T\right)+O\left(N^2T^2\right) & O\left(N^3T\right)+O\left(NT^2\right)\\ O\left(NT^3\right)+O\left(N^2T\right) & O\left(N^3T\right)+O\left(NT^2\right) & O\left(N^2T\right)+O\left(NT^2\right)\end{pmatrix},\end{aligned}$$

which, together with (23), implies the proposition.

Inverse of $\hat{\Omega}_\eta$**.** Let

$$\begin{aligned}\Omega_\eta &= A^{*(D)}\left(I_N\otimes J_T\otimes\tilde{\Sigma}_\lambda+J_N\otimes I_T\otimes\tilde{\Sigma}_\tau+I_{NT}\otimes\tilde{\Sigma}_\xi\right)A^{*(D)\prime}\\ &= A^{*(B)}\left(I_N\otimes\tilde{\Sigma}_\lambda\right)A^{*(B)\prime}+A^{*(T)}\left(I_T\otimes\tilde{\Sigma}_\tau\right)A^{*(T)\prime}+A^{*(D)}\left(I_{NT}\otimes\tilde{\Sigma}_\xi\right)A^{*(D)\prime},\end{aligned}$$

then we have, by the equality

$$\left(U+XVX'\right)^{-1} = U^{-1}-U^{-1}X\left(V^{-1}+X'U^{-1}X\right)^{-1}X'U^{-1} \tag{24}$$

from Searle et al. (1992) p. 453, that

$$\Omega_\eta^{-1} = C^{-1}-C^{-1}A^{*(T)}\left\{A^{*(T)\prime}C^{-1}A^{*(T)}+\left(I_T\otimes\tilde{\Sigma}_\tau^{-1}\right)\right\}^{-1}A^{*(T)\prime}C^{-1} \tag{25}$$

with

$$\begin{aligned} C^{-1} &= \left[A^{*(D)}\left(I_{NT}\otimes\tilde{\Sigma}_{\xi}\right)A^{*(D)\prime}+A^{*(B)}\left(I_{N}\otimes\tilde{\Sigma}_{\lambda}\right)A^{*(B)\prime}\right]^{-1} \\ &= D^{-1}-D^{-1}A^{*(B)}\left\{A^{*(B)\prime}D^{-1}A^{*(B)}+\left(I_{N}\otimes\tilde{\Sigma}_{\lambda}^{-1}\right)\right\}^{-1}A^{*(B)\prime}D^{-1} \end{aligned} \tag{26}$$

and

$$\begin{aligned} D^{-1} &= \left[A^{*(D)}\left(I_{NT}\otimes\tilde{\Sigma}_{\xi}\right)A^{*(D)\prime}\right]^{-1} \\ &= \mathrm{Diag}_{i=1}^{N}\left[D_{i}^{-1}\right], \end{aligned} \tag{27}$$

where $D_j^{-1}=\mathrm{Diag}_{t=1}^{T}\left[d_{jt}^{-1}\right]$ and $d_{jt}=a_{jt}^{*\prime}\tilde{\Sigma}_{\xi}a_{jt}^{*}$.

B.3 Feasible Aitken Estimation

Let $\mathbf{D}_{LM}\left(N^{l}T^{m};\Sigma\right)$ denote a $LM\times LM$ diagonal matrix with nonzero elements that are of order N^lT^m and that are a function of the matrix Σ. Likewise, $\mathbf{S}_{LM}\left(N^{l}T^{m};\Sigma\right)$ denotes a $LM\times LM$ matrix , $\mathbf{B}_{LM;M}\left(N^{l}T^{m};\Sigma\right)$ a $LM\times LM$ block-diagonal matrix with $M\times M$ blocks and $\mathbf{T}_{LM;M}\left(N^{l}T^{m};\Sigma\right)$ a $LM\times LM$ matrix with nonzero elements where $J_L\otimes I_M$ has nonzero elements. The second argument of these matrices denotes that their nonzero elements are functions of the matrix Σ.

The proof of Theorem 3 consists of two parts. The first part proves the equivalence of Aitken and feasible Aitken estimators in the case we are dealing with an error-components model. The second part proves the rate of convergence of the variance in the general case. It makes use of the following lemma.

Lemma D.1. Under Assumption E, the inverse of the variance matrix of the disturbances, Ω_{η}^{-1}, can be asymptotically approximated by $\Phi_A+\Phi_B+\Phi_C+\Phi_D$ in the sense that, for conformable matrices X_1 and X_2, it holds that

$$X_1'\Omega_{\eta}^{-1}X_2 \quad \overset{p}{\longrightarrow} \quad X_1'\left(\Phi_A+\Phi_B+\Phi_C+\Phi_D\right)X_2 \tag{28}$$

where

$$\begin{array}{rcccl} \Phi_A &=& D^{-1} &=& \mathbf{D}_{NT}\left(1;\tilde{\Sigma}_{\xi}\right) \\ \Phi_B &=& -D^{-1}A^{*(B)}\mathrm{Diag}_{i=1}^{N}\left[E_i^{-1}\right]A^{*(B)\prime}D^{-1} &=& \mathbf{B}_{NT;T}\left(T^{-1};\tilde{\Sigma}_{\xi}\right) \\ \Phi_C &=& D^{-1}A^{*(T)}G_1^{-1}A^{*(T)\prime}D^{-1} &=& \mathbf{T}_{NT;T}\left(N^{-1};\tilde{\Sigma}_{\xi}\right) \\ \Phi_D &=& D^{-1}A^{*(T)}G_1^{-1}G_2G_1^{-1}A^{*(T)\prime}D^{-1} &=& \mathbf{S}_{NT}\left(N^{-1}T^{-1};\tilde{\Sigma}_{\xi}\right) \end{array}$$

with

$$\begin{aligned} E_i &= \sum_{s=1}^{T} \left(a_{is}^{*\prime}\tilde{\Sigma}_\xi w_{is}\right)^{-1} a_{is}^{*} a_{is}^{*\prime} \\ G_1 &= A^{*(T)\prime} D^{-1} A^{*(T)} \\ G_2 &= A^{*(T)\prime} D^{-1} A^{*(B)} \mathrm{Diag}_{i=1}^{N}\left[E_i^{-1}\right] A^{*(B)\prime} D^{-1} A^{*(T)}. \end{aligned}$$

Proof. The inverse of Ω_η^{-1} is given by (25). Using the above notation and (27), the inverse of the $KN \times KN$ matrix $A^{*(B)\prime} D^{-1} A^{*(B)} + I_N \otimes \tilde{\Sigma}_\lambda^{-1}$ can be written as

$$\begin{aligned} & \mathrm{Diag}_{i=1}^{N}\left[\left(\sum_{s=1}^{T} a_{is}^{*}\left(a_{is}^{*\prime}\tilde{\Sigma}_\xi w_{is}\right)^{-1} a_{is}^{*\prime} + \tilde{\Sigma}_\lambda^{-1}\right)^{-1}\right] \\ \approx\ & \mathrm{Diag}_{i=1}^{N}\left[E_i^{-1}\right] \\ =\ & \mathbf{B}_{NK;K}\left(T^{-1}; \tilde{\Sigma}_\xi\right) \end{aligned}$$

since

$$\begin{aligned} \left(E_i + \tilde{\Sigma}_\lambda^{-1}\right)^{-1} &= E_i^{-1} - E_i^{-1}\left(\tilde{\Sigma}_\lambda + E_i^{-1}\right)^{-1} E_i^{-1} \\ &= E_i^{-1} - E_i^{-1}\tilde{\Sigma}_\lambda^{-1} E_i^{-1} + E_i^{-1}\tilde{\Sigma}_\lambda^{-1}\left(\tilde{\Sigma}_\lambda^{-1} + E_i\right)^{-1}\tilde{\Sigma}_\lambda^{-1} E_i^{-1}, \end{aligned}$$

with $E_i = \sum_{s=1}^{T} a_{is}^{*}\left(a_{is}^{*\prime}\tilde{\Sigma}_\xi a_{is}^{*}\right)^{-1} a_{is}^{*\prime} = \mathbf{S}_K\left(T; \tilde{\Sigma}_\xi\right)$. Now it holds that $E_i^{-1} = \mathbf{S}_K\left(T^{-1}; \tilde{\Sigma}_\xi\right)$, that $E_i^{-1}\tilde{\Sigma}_\lambda^{-1} E_i^{-1} = \mathbf{S}_K\left(T^{-2}; \tilde{\Sigma}_\lambda, \tilde{\Sigma}_\xi\right)$ and that $E_i^{-1}\tilde{\Sigma}_\lambda^{-1}\left(\tilde{\Sigma}_\lambda^{-1} + E_i\right)^{-1}\tilde{\Sigma}_\lambda^{-1} E_i^{-1} = \mathbf{S}_K\left(T^{-3}; \tilde{\Sigma}_\lambda, \tilde{\Sigma}_\xi\right)$, from which we have that

$$X_1' \mathrm{Diag}_{i=1}^{N}\left[\left(E_i + \tilde{\Sigma}_\lambda^{-1}\right)^{-1}\right] X_2 \quad \overset{p}{\longrightarrow} \quad X_1' \mathrm{Diag}_{i=1}^{N}\left[E_i^{-1}\right] X_2,$$

for which the notation $\mathrm{Diag}_{i=1}^{N}\left[\left(E_i + \tilde{\Sigma}_\lambda^{-1}\right)^{-1}\right] \approx \mathrm{Diag}_{i=1}^{N}\left[E_i^{-1}\right]$ is used. Consequently, it holds that

$$\begin{aligned} C^{-1} &= D^{-1} - D^{-1} A^{*(B)}\left[A^{*(B)\prime} D^{-1} A^{*(B)} + I_N \otimes \tilde{\Sigma}_\lambda^{-1}\right]^{-1} A^{*(B)\prime} D^{-1} \\ &\approx D^{-1} - D^{-1} A^{*(B)} \mathrm{Diag}_{i=1}^{N}\left[E_i^{-1}\right] A^{*(B)\prime} D^{-1} \\ &= \mathbf{D}_{NT}\left(1; \tilde{\Sigma}_\xi\right) - \mathbf{B}_{NT;T}\left(T^{-1}; \tilde{\Sigma}_\xi\right). \end{aligned}$$

Define now $F = A^{*(T)\prime} C^{-1} A^{*(T)} + \left(I_T \otimes \tilde{\Sigma}_\tau^{-1}\right)$, then we have that

$$F \approx G_1 - G_2 + G_3,$$

where

$$\begin{array}{rcccl} G_1 & = & A^{*(T)\prime}D^{-1}A^{*(T)} & = & \mathbf{B}_{TK;K}\left(N;\tilde{\Sigma}_\xi\right) \\ G_2 & = & A^{*(T)\prime}D^{-1}A^{*(B)}\mathrm{Diag}_{i=1}^{N}\left[E_i^{-1}\right]A^{*(B)\prime}D^{-1}A^{*(T)} & = & \mathbf{S}_{TK}\left(NT^{-1};\tilde{\Sigma}_\xi\right) \\ G_3 & = & \left(I_T\otimes\tilde{\Sigma}_\tau^{-1}\right) & = & \mathbf{B}_{TK;K}\left(1;\tilde{\Sigma}_\tau\right) \end{array}$$

and

$$\begin{aligned} F^{-1} &\approx (G_1-G_2+G_3)^{-1} \\ &= G_1^{-1}+G_1^{-1}G_2G_1^{-1}-G_1^{-1}G_3G_1^{-1} \\ &\approx G_1^{-1}+G_1^{-1}G_2G_1^{-1} \\ &= \mathbf{B}_{TK;K}\left(N^{-1};\tilde{\Sigma}_\xi\right)+\mathbf{S}_{TK}\left(N^{-1}T^{-1};\tilde{\Sigma}_\xi\right), \end{aligned}$$

by applying (24) and because $G_1^{-1}G_3G_1^{-1}=\mathbf{B}_{TK;K}\left(N^{-2};\tilde{\Sigma}_\tau,\tilde{\Sigma}_\xi\right)$. Consequently, it holds that

$$\begin{aligned} C^{-1}A^{*(T)}F^{-1}A^{*(T)\prime}C^{-1} &\approx C^{-1}A^{*(T)}G_1^{-1}A^{*(T)\prime}C^{-1} \\ &\quad +C^{-1}A^{*(T)}G_1^{-1}G_2G_1^{-1}A^{*(T)\prime}C^{-1} \\ &= \mathbf{T}_{NT;T}\left(N^{-1};\tilde{\Sigma}_\xi\right)+\mathbf{S}_{NT}\left(N^{-1}T^{-1};\tilde{\Sigma}_\xi\right) \end{aligned}$$

and the inverse covariance matrix can be written as

$$\begin{aligned} \Omega_\eta^{-1} &= C^{-1}-C^{-1}A^{*(T)}F^{-1}A^{*(T)\prime}C^{-1} \\ &\approx D^{-1}-D^{-1}A^{*(B)}\mathrm{Diag}_{i=1}^{N}\left[E_i^{-1}\right]A^{*(B)\prime}D^{-1} \\ &\quad +D^{-1}A^{*(T)}G_1^{-1}A^{*(T)\prime}D^{-1} \\ &\quad +D^{-1}A^{*(T)}G_1^{-1}G_2G_1^{-1}A^{*(T)\prime}D^{-1} \\ &= \mathbf{D}_{NT}\left(1;\tilde{\Sigma}_\xi\right)-\mathbf{B}_{NT;T}\left(T^{-1};\tilde{\Sigma}_\xi\right) \\ &\quad +\mathbf{T}_{NT;T}\left(N^{-1};\tilde{\Sigma}_\xi\right)+\mathbf{S}_{NT}\left(N^{-1}T^{-1};\tilde{\Sigma}_\xi\right), \end{aligned}$$

which proves the lemma.

Remark. Note that $X_1\Phi_JX_2=O\left(NT\right)$ for $J=A,B,C,D$, for all conformable matrices $X_1=O_p\left(1\right)$ and $X_2=O_p\left(1\right)$. Note also that the leading terms in the asymptotic approximation of Ω_η^{-1} are all functions of $\tilde{\Sigma}_\xi$, but neither of $\tilde{\Sigma}_\lambda$ nor $\tilde{\Sigma}_\tau$.

Proof of Theorem 3. To prove the first part of the Theorem, we need to prove that

$$\begin{aligned} \mathrm{plim}_{N,T\to\infty}\left[\left(NT\right)^{-1}\hat{R}'\left(\hat{\Omega}_\eta^{-1}-\Omega_\eta^{-1}\right)R\right] &= 0 \\ \mathrm{plim}_{N,T\to\infty}\left[\left(NT\right)^{-1}\Psi_\Omega\hat{R}'\left(\hat{\Omega}_\eta^{-1}-\Omega_\eta^{-1}\right)H\right] &= 0. \end{aligned}$$

We have, by the consistency of $\left(\hat{\sigma}_\lambda^2, \hat{\sigma}_\tau^2, \hat{\sigma}_\xi^2\right)$, that

$$\begin{aligned}
\text{plim}_{N,T\to\infty}\left[(NT)^{-1}\hat{R}'\left(\hat{\Omega}_\eta^{-1}-\Omega_\eta^{-1}\right)R\right] &= \text{plim}_{N,T\to\infty}\left[(NT)^{-1}\hat{R}'\mathcal{P}_i R\left(\frac{\sigma_\lambda^2-\hat{\sigma}_\lambda^2}{\hat{\sigma}_\lambda^2\sigma_\lambda^2}\right)\right] \\
&\quad +\text{plim}_{N,T\to\infty}\left[(NT)^{-1}\hat{R}'\mathcal{P}_p R\left(\frac{\sigma_\tau^2-\hat{\sigma}_\tau^2}{\hat{\sigma}_\tau^2\sigma_\tau^2}\right)\right] \\
&\quad +\text{plim}_{N,T\to\infty}\left[(NT)^{-1}\hat{R}'\mathcal{P}_r R\left(\frac{\sigma_\xi^2-\hat{\sigma}_\xi^2}{\hat{\sigma}_\xi^2\sigma_\xi^2}\right)\right] \\
&= o(1),
\end{aligned}$$

and that

$$\begin{aligned}
(NT)^{-1}\Psi_\Omega\hat{R}'\hat{\Omega}_\eta^{-1}H &= (NT)^{-1}\left(T\hat{\sigma}_\lambda^2\right)^{-1}\Psi_\Omega\hat{R}'\mathcal{P}_i H \\
&\quad +(NT)^{-1}\left(N\hat{\sigma}_\tau^2\right)^{-1}\Psi_\Omega\hat{R}'\mathcal{P}_p H \\
&\quad +(NT)^{-1}\hat{\sigma}_\xi^{-2}\Psi_\Omega\hat{R}'\mathcal{P}_r H \\
&= T_1+T_2+T_3+T_4,
\end{aligned}$$

where

$$\begin{aligned}
T_1 &= (NT)^{-1}\Psi_\Omega\hat{R}'\Omega_\eta^{-1}H \\
T_2 &= N^{-1}T^{-2}\sigma_\lambda^{-4}\left(\sigma_\lambda^2-\hat{\sigma}_\lambda^2\right)\Psi_\Omega\hat{R}'\mathcal{P}_i H \\
T_3 &= N^{-2}T^{-1}\sigma_\tau^{-4}\left(\sigma_\tau^2-\hat{\sigma}_\tau^2\right)\Psi_\Omega\hat{R}'\mathcal{P}_p H \\
T_4 &= (NT)^{-1}\sigma_\xi^{-4}\left(\sigma_\xi^2-\hat{\sigma}_\xi^2\right)\Psi_\Omega\hat{R}'\mathcal{P}_r H
\end{aligned}$$

The first term, T_1, is the numerator of the Aitken estimator (see Theorem 2) and it has variance equal to $O(1)$. The second term can be written as

$$\begin{aligned}
T_2 &= N^{-1}T^{-2}\sigma_\lambda^{-4}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\Psi_\Omega\hat{r}_{it}\left(\lambda_i+\frac{\xi_{it}}{T}\right)\right. \\
&\quad \left.\times\frac{2}{NT(T-1)}\sum_{j=1}^{N}\sum_{r=1}^{T}\sum_{s=r+1}^{T}\left(\eta_{jr}\eta_{js}-\mathrm{E}\left[\eta_{jr}\eta_{js}\right]\right)\right\},
\end{aligned}$$

where the second line is equal to $\sigma_\lambda^2-\hat{\sigma}_\lambda^2$, with $\hat{\sigma}_\lambda^2$ the covariance estimator from subsection 4.2 in the case of an error component model. T_2 has variance equal to

$$\text{Var}\left[T_2\right] = \frac{4\sigma_\lambda^{-8}}{N^4T^6(T-1)^2}\sum_{i,j,k,l}^{N}\sum_{s,t,u,w}^{T}\sum_{v>u,x>w}^{T}\Psi_\Omega\mathrm{E}\left[\hat{r}_{it}\left(\lambda_i+\frac{\xi_{it}}{T}\right)\left(\eta_{ku}\eta_{kv}-\mathrm{E}\left[\eta_{ku}\eta_{kv}\right]\right)\right.$$

$$
\begin{aligned}
&\times \left(\eta_{lw}\eta_{lx} - \mathrm{E}\left[\eta_{lw}\eta_{lx}\right]\right)\left(\lambda_j + \frac{\xi_{js}}{T}\right)\hat{r}'_{js}\Big]\Psi_\Omega \\
&-\frac{4\sigma_\lambda^{-8}}{N^4T^6\left(T-1\right)^2}\left(\sum_{i,k}^{N}\sum_{t,u}^{T}\sum_{v>u}^{T}\Psi_\Omega \mathrm{E}\left[\hat{r}_{it}\left(\lambda_i + \frac{\xi_{it}}{T}\right)\left(\eta_{ku}\eta_{kv} - \mathrm{E}\left[\eta_{ku}\eta_{kv}\right]\right)\right]\right)^2 \\
= \;& O\left(N^{-2}T^{-2}\right)\Psi_\Omega J_{(1+K)\left(1+\frac{K}{2}+J\right)}\Psi_\Omega,
\end{aligned}
$$

since

$$
\begin{aligned}
&\sum_{i,j,k,l}^{N}\sum_{s,t,u,v,w,x}^{T}\mathrm{E}\left[\left(\eta_{ku}\eta_{kv} - \mathrm{E}\left[\eta_{ku}\eta_{kv}\right]\right)\left(\lambda_i + \frac{\xi_{it}}{T}\right)\left(\lambda_j + \frac{\xi_{js}}{T}\right)\left(\eta_{lw}\eta_{lx} - \mathrm{E}\left[\eta_{lw}\eta_{lx}\right]\right)\right] \\
- \;& \left(\sum_{i,k}^{N}\sum_{t,u,v}^{T}\mathrm{E}\left[\left(\eta_{ku}\eta_{kv} - \mathrm{E}\left[\eta_{ku}\eta_{kv}\right]\right)\left(\lambda_i + \frac{\xi_{it}}{T}\right)\right]\right)^2 \\
= \;& O\left(N^2T^6\right).
\end{aligned}
$$

Similarly we have that $\mathrm{Var}\left[T_3\right] = O\left(N^{-2}T^{-2}\right)\Psi_\Omega J_{(1+K)\left(1+\frac{K}{2}+J\right)}\Psi_\Omega$ and that

$$
\begin{aligned}
\mathrm{Var}\left[T_4\right] = \;& \frac{\sigma_\xi^{-8}}{N^4T^4}\sum_{i,j,k,l}^{N}\sum_{s,t,u,v}^{T}\mathrm{E}\Big[\hat{r}_{it}\left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right)\xi_{it} \\
&\times\xi_{js}\left(\eta_{lv}^2 - \mathrm{E}\left[\eta_{lv}^2\right]\right)\hat{r}'_{js}\Big] \\
&-\frac{\sigma_\xi^{-8}}{N^4T^4}\left(\sum_{i,k}^{N}\sum_{t,u}^{T}\mathrm{E}\left[\hat{r}_{it}\left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right)\xi_{it}\right]\right)^2 \\
= \;& O\left(N^{-2}T^{-2}\right)\Psi_\Omega J_{(1+K)\left(1+\frac{K}{2}+J\right)}\Psi_\Omega.
\end{aligned}
$$

Since we have, from Theorem 2, that $\Psi_\Omega = \mathrm{Diag}\left[\sqrt{N}I_{\dim R_i}, \sqrt{T}I_{\dim R_p}, \sqrt{NT}I_{\dim R_r}\right]$, the above expressions prove that

$$
\begin{aligned}
\left(NT\right)^{-1}\hat{R}'\hat{\Omega}_\eta^{-1}H \quad &\overset{p}{\longrightarrow} \quad 0, \\
\left(NT\right)^{-1}\Psi_\Omega\hat{R}'\hat{\Omega}_\eta^{-1}H \quad &\overset{d}{\longrightarrow} \quad \left(NT\right)^{-1}\Psi_\Omega\hat{R}'\Omega_\eta^{-1}H,
\end{aligned}
$$

and thus the first part of the theorem.

To prove the second part, note that

$$
\begin{aligned}
\left(NT\right)^{-1}\hat{R}'\hat{\Omega}_\eta^{-1}H \quad \overset{p}{\longrightarrow} \quad & \left(NT\right)^{-1}\hat{R}'\hat{D}^{-1}Z \\
& - \left(NT\right)^{-1}\hat{R}'\hat{D}^{-1}A^{*(B)}\mathrm{Diag}_{i=1}^{N}\left[\hat{E}_i^{-1}\right]A^{*(B)\prime}\hat{D}^{-1}Z
\end{aligned}
$$

$$+ (NT)^{-1} \hat{R}' \hat{D}^{-1} A^{*(T)} \hat{G}_1^{-1} A^{*(T)\prime} \hat{D}^{-1} Z$$
$$+ (NT)^{-1} \hat{R}' \hat{D}^{-1} A^{*(T)} \hat{G}_1^{-1} \hat{H}_1 \hat{G}_1^{-1} A^{*(T)\prime} \hat{D}^{-1} Z, \qquad (29)$$

with $Z = R, H$, by lemma D.1.

First it holds that $(NT)^{-1} \hat{R}' \hat{\Omega}_\eta^{-1} R = O_p(1)$, because $(NT)^{-1} \hat{R}' \hat{D}^{-1} R$ has as p,q-th element

$$\begin{aligned}(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{r}_{it;p} \left(a_{it}^{*\prime} \hat{\tilde{\Sigma}}_\xi a_{it}^* \right)^{-1} r_{it;q} &= (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{r}_{it;p} \left(a_{it}^{*\prime} \tilde{\Sigma}_\xi a_{it}^* \right)^{-1} r_{it;q} \\ &+ (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{r}_{it;q} \left\{ \left(a_{it}^{*\prime} \tilde{\Sigma}_\xi a_{it}^* \right)^{-2} w_{it}^{(\xi)\prime} \right. \\ &\quad \left. \times \left[W^{(\xi)\prime} W^{(\xi)} \right]^{-1} \left[W^{(\xi)\prime} E^{(\xi)} \right] \right. \\ &\quad \left. + O_p(\mathbf{R}_r) \right\} r_{it;q},\end{aligned}$$

with $O_p(\mathbf{R}_r) = O_p\left(\sum_{j,k}^{k_\eta} \left(\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)} \right)_j \left(\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)} \right)_k \right)$, since

$$\begin{aligned}\left(a_{it}^{*\prime} \hat{\tilde{\Sigma}}_\xi a_{it}^* \right)^{-1} &= \left(a_{it}^{*\prime} \tilde{\Sigma}_\xi a_{it}^* \right)^{-1} + \left(a_{it}^{*\prime} \tilde{\Sigma}_\xi a_{it}^* \right)^{-2} w_{it}^{(\xi)\prime} \left(\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)} \right) \\ &\quad + O_p\left(\sum_{j,k}^{k_\eta} \left(\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)} \right)_j \left(\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)} \right)_k \right),\end{aligned}$$

by (15), using the definitions of $\tilde{\Sigma}^{(\xi)}$ and $w_{it}^{(\xi)}$ from subsection 4.2. We further have that $(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{r}_{it;p} \left(a_{it}^{*\prime} \tilde{\Sigma}_\xi a_{it}^* \right)^{-1} r_{it;q} = O_p(1)$, and that

$$(NT)^{-1} \sum_{i,j=1}^{N} \sum_{s,t=1}^{T} \left\{ \hat{r}_{it;p} \left(a_{it}^{*\prime} \tilde{\Sigma}_\xi a_{it}^* \right)^{-2} w_{it}^{(\xi)\prime} \times \left[W^{(\xi)\prime} W^{(\xi)} \right]^{-1} w_{js}^{(\xi)} \left(\eta_{js}^2 - \mathrm{E}\left[\eta_{js}^2 \right] \right) r_{it;q} \right. = o_p(1),$$

which proves $(NT)^{-1} \hat{R}' \hat{D}^{-1} R = O_p(1)$. The other terms in (29) are of of the same order in probability as $(NT)^{-1} \hat{R}' \hat{D}^{-1} R$ and hence $(NT)^{-1} \hat{R}' \hat{\Omega}_\eta^{-1} R = O_p(1)$.

Secondly, we have that $(NT)^{-1} \hat{R}' \hat{D}^{-1} H$ has as q^{th} element

$$\begin{aligned}(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{r}_{it;q} \left(a_{it}^{*\prime} \hat{\tilde{\Sigma}}_\xi a_{it}^* \right)^{-1} \eta_{it} &= (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{r}_{it;q} \left(a_{it}^{*\prime} \tilde{\Sigma}_\xi a_{it}^* \right)^{-1} \eta_{it} \qquad (30) \\ &+ (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{r}_{it;q} \left\{ \left(a_{it}^{*\prime} \tilde{\Sigma}_\xi a_{it}^* \right)^{-2} w_{it}^{(\xi)\prime} \right. \\ &\quad \left. \times \left[W^{(\xi)\prime} W^{(\xi)} \right]^{-1} \left[W^{(\xi)\prime} E^{(\xi)} \right] \right.\end{aligned}$$

$$
\begin{aligned}
&+\left(a_{it}^{*\prime}\tilde{\Sigma}_{\xi}a_{it}^{*}\right)^{-3}w_{it}^{(\xi)\prime}\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}\\
&\quad\times\left[W^{(\xi)\prime}E^{(\xi)}\right]\left[E^{(\xi)\prime}W^{(\xi)}\right]\\
&\quad\times\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}w_{it}^{(\xi)}\\
&+O_{p}\left(\mathbf{R}_{\eta}\right)\}\,\eta_{it},
\end{aligned}
$$

with $O_{p}\left(\mathbf{R}_{\eta}\right)=O_{p}\left(\sum_{j,k,l}^{k_{\eta}}\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_{j}\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_{k}\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_{l}\right)$, since

$$
\begin{aligned}
\left(a_{it}^{*\prime}\hat{\tilde{\Sigma}}_{\xi}a_{it}^{*}\right)^{-1} &= \left(a_{it}^{*\prime}\tilde{\Sigma}_{\xi}a_{it}^{*}\right)^{-1}+\left(a_{it}^{*\prime}\tilde{\Sigma}_{\xi}a_{it}^{*}\right)^{-2}w_{it}^{(\xi)\prime}\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)\\
&\quad+\left(a_{it}^{*\prime}\tilde{\Sigma}_{\xi}a_{it}^{*}\right)^{-3}w_{it}^{(\xi)\prime}\left(\left[\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right]\left[\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right]^{\prime}\right)w_{it}^{(\xi)}\\
&\quad+O_{p}\left(\sum_{j,k,l}^{k_{\eta}}\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_{j}\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_{k}\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_{l}\right),
\end{aligned}
$$

by (15), using the definitions of $\tilde{\Sigma}^{(\xi)}$ and $w_{it}^{(\xi)}$ from subsection 4.2. The first term of (30), $T_{1}=(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;q}\left(a_{it}^{*\prime}\tilde{\Sigma}_{\xi}a_{it}^{*}\right)^{-1}\eta_{it}$, has variance equal to $O_{p}\left((NT)^{-1}\right)$. The second term can be written as

$$
\begin{aligned}
T_{2} &= (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\hat{r}_{it;q}\left(a_{it}^{*\prime}\tilde{\Sigma}_{\xi}a_{it}^{*}\right)^{-2}w_{it}^{(\xi)\prime}\right.\\
&\quad\left.\times\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T}w_{js}^{(\xi)}\left(\eta_{js}^{2}-\mathrm{E}\left[\eta_{js}^{2}\right]\right)\eta_{it}\right\},
\end{aligned}
$$

and its variance has q^{th} diagonal element equal to

$$
\begin{aligned}
\mathrm{Var}\left[T_{2}\right] &= (NT)^{-2}\sum_{i,j,k,l}^{N}\sum_{s,t,u,v}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(a_{it}^{*\prime}\tilde{\Sigma}_{\xi}a_{it}^{*}\right)^{-2}w_{it}^{(\xi)\prime}\right.\\
&\quad\times\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}w_{ku}^{(\xi)}\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it}\\
&\quad\times\eta_{js}\left(\eta_{lv}^{2}-\mathrm{E}\left[\eta_{lv}^{2}\right]\right)w_{lv}^{(\xi)\prime}\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}\\
&\quad\left.\times w_{ku}^{(\xi)}\left(a_{js}^{*\prime}\tilde{\Sigma}_{\xi}a_{js}^{*}\right)^{-2}\hat{r}_{js;q}\right]\\
&\quad-(NT)^{-2}\left(\sum_{i,k}^{N}\sum_{t,u}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(a_{it}^{*\prime}\tilde{\Sigma}_{\xi}a_{it}^{*}\right)^{-2}w_{it}^{(\xi)\prime}\right.\right.\\
&\quad\left.\left.\times\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}w_{ku}^{(\xi)}\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it}\right]\right)^{2}
\end{aligned}
$$

$$= O\left(N^{-2}+T^{-2}\right),$$

since

$$\begin{aligned}
&\sum_{i,j,k,l}^{N}\sum_{s,t,u,v}^{T}\mathrm{E}\left[\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it}\eta_{js}\left(\eta_{lv}^{2}-\mathrm{E}\left[\eta_{lv}^{2}\right]\right)\right] \\
&-\left(\sum_{i,k}^{N}\sum_{t,u}^{T}\mathrm{E}\left[\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it}\right]\right)^{2} \\
= \;& O\left(N^{2}T^{4}+N^{4}T^{2}\right), \qquad (31)
\end{aligned}$$

and since $W^{(\xi)\prime}W^{(\xi)}=O\left(NT\right)$. Similarly, the q,r-th element of $\mathrm{Cov}\left[T_1T_2\right]$ is given by

$$\begin{aligned}
\mathrm{Cov}\left[T_2T_1\right] &= (NT)^{-2}\sum_{i,j,k}^{N}\sum_{s,t,u}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(a_{it}^{*\prime}\tilde{\Sigma}_{\xi}a_{it}^{*}\right)^{-2}w_{it}^{(\xi)\prime}\right. \\
&\quad\times\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}w_{ku}^{(\xi)}\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it} \\
&\quad\left.\times\eta_{js}\left(a_{js}^{*\prime}\tilde{\Sigma}_{\xi}a_{js}^{*}\right)^{-1}\hat{r}_{js;r}\right] \\
&= O\left(N^{-2}+T^{-2}\right),
\end{aligned}$$

since

$$\sum_{i,j,k}^{N}\sum_{s,t,u}^{T}\mathrm{E}\left[\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it}\eta_{js}\right] = O\left(NT^{3}+N^{3}T\right).$$

T_3 can be written as

$$\begin{aligned}
T_3 &= (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\hat{r}_{it;q}\left(a_{it}^{*\prime}\tilde{\Sigma}_{\xi}a_{it}^{*}\right)^{-3}w_{it}^{(\xi)\prime}\right. \\
&\quad\times\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T}w_{js}^{(\xi)}\left(\eta_{js}^{2}-\mathrm{E}\left[\eta_{js}^{2}\right]\right) \\
&\quad\left.\times\sum_{k=1}^{N}\sum_{u=1}^{T}\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)w_{ku}^{(\xi)\prime}\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}w_{it}^{(\xi)}\eta_{it}\right\},
\end{aligned}$$

with q^{th} diagonal element of its variance

$$\begin{aligned}
\mathrm{Var}\left[T_3\right] &= (NT)^{-2}\sum_{i,j,k,l,m,n}^{N}\sum_{s,t,u,v,x,y}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(a_{it}^{*\prime}\tilde{\Sigma}_{\xi}a_{it}^{*}\right)^{-3}w_{it}^{(\xi)\prime}\right. \\
&\quad\left.\times\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}w_{ku}^{(\xi)}\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\right.
\end{aligned}$$

$$
\begin{aligned}
&\times\left(\eta_{lv}^2-\mathrm{E}\left[\eta_{lv}^2\right]\right)w_{lv}^{(\xi)\prime}\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}\\
&\times w_{it}^{(\xi)}\eta_{it}\eta_{js}w_{js}^{(\xi)\prime}\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}\\
&\times w_{mx}^{(\xi)}\left(\eta_{mx}^2-\mathrm{E}\left[\eta_{mx}^2\right]\right)\left(\eta_{ny}^2-\mathrm{E}\left[\eta_{ny}^2\right]\right)w_{ly}^{(\xi)\prime}\\
&\times\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}w_{js}^{(\xi)}\left(a_{js}^{*\prime}\tilde{\Sigma}_\xi a_{js}^{*}\right)^{-3}\hat{r}_{js;q}\Bigg]\\
-(NT)^{-2}&\left(\sum_{i,j,k}^{N}\sum_{s,t,u}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(a_{it}^{*\prime}\tilde{\Sigma}_\xi a_{it}^{*}\right)^{-3}w_{it}^{(\xi)\prime}\right.\right.\\
&\times\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}w_{js}^{(\xi)}\left(\eta_{js}^2-\mathrm{E}\left[\eta_{js}^2\right]\right)\\
&\left.\left.\times\left(\eta_{ku}^2-\mathrm{E}\left[\eta_{ku}^2\right]\right)w_{ku}^{(\xi)\prime}\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}w_{it}^{(\xi)}\eta_{it}\right]\right)^2\\
=\;&O\left(N^{-3}+T^{-3}\right),
\end{aligned}
$$

since

$$
\begin{aligned}
&\sum_{i,j,k,l,m,n}^{N}\sum_{s,t,u,v,x,y}^{T}\mathrm{E}\left[\left(\eta_{ku}^2-\mathrm{E}\left[\eta_{ku}^2\right]\right)\left(\eta_{lv}^2-\mathrm{E}\left[\eta_{lv}^2\right]\right)\right.\\
&\qquad\left.\times\eta_{it}\eta_{js}\left(\eta_{mx}^2-\mathrm{E}\left[\eta_{mx}^2\right]\right)\left(\eta_{ny}^2-\mathrm{E}\left[\eta_{ny}^2\right]\right)\right]\\
&-\left(\sum_{i,j,k}^{N}\sum_{s,t,u}^{T}\mathrm{E}\left[\left(\eta_{js}^2-\mathrm{E}\left[\eta_{js}^2\right]\right)\left(\eta_{ku}^2-\mathrm{E}\left[\eta_{ku}^2\right]\right)\eta_{it}\right]\right)^2\\
=\;&O\left(N^3T^6+N^6T^3\right).
\end{aligned}
$$

The q,r-th element of $\mathrm{Cov}\left[T_3T_1\right]$ is given by

$$
\begin{aligned}
\mathrm{Cov}\left[T_3,T_1\right]\;=\;&(NT)^{-2}\sum_{i,j,k,l}^{N}\sum_{s,t,u,v}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(a_{it}^{*\prime}\tilde{\Sigma}_\xi a_{it}^{*}\right)^{-3}w_{it}^{(\xi)\prime}\right.\\
&\times\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}w_{ku}^{(\xi)}\left(\eta_{ku}^2-\mathrm{E}\left[\eta_{ku}^2\right]\right)\\
&\times\left(\eta_{lv}^2-\mathrm{E}\left[\eta_{lv}^2\right]\right)w_{lv}^{(\xi)\prime}\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}\\
&\left.\times w_{it}^{(\xi)}\eta_{it}\eta_{js}\left(a_{js}^{*\prime}\tilde{\Sigma}_\xi a_{js}^{*}\right)^{-1}\hat{r}_{js;r}\right]\\
=\;&O\left(N^{-2}+T^{-2}\right),
\end{aligned}
$$

by (31), and the q,r-th element of $\mathrm{Cov}\left[T_3T_2\right]$ is given by

$$
\begin{aligned}
\mathrm{Cov}\left[T_3,T_2\right]\;=\;&(NT)^{-2}\sum_{i,j,k,l,m}^{N}\sum_{s,t,u,v,x}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(a_{it}^{*\prime}\tilde{\Sigma}_\xi a_{it}^{*}\right)^{-3}w_{it}^{(\xi)\prime}\right.\\
&\left.\times\left[W^{(\xi)\prime}W^{(\xi)}\right]^{-1}w_{ku}^{(\xi)}\left(\eta_{ku}^2-\mathrm{E}\left[\eta_{ku}^2\right]\right)\right.
\end{aligned}
$$

$$
\begin{aligned}
&\times \left(\eta_{lv}^2 - \mathrm{E}\left[\eta_{lv}^2\right]\right) w_{lv}^{(\xi)\prime} \left[W^{(\xi)\prime} W^{(\xi)}\right]^{-1} \\
&\times w_{it}^{(\xi)} \eta_{it}\eta_{js} \left(\eta_{mx}^2 - \mathrm{E}\left[\eta_{mx}^2\right]\right) w_{mx}^{(\xi)} \\
&\times \left[W^{(\xi)\prime} W^{(\xi)}\right]^{-1} w_{js}^{(\xi)} \left(a_{js}^{*\prime} \tilde{\Sigma}_\xi a_{js}^*\right)^{-2} \hat{r}_{js;q}\Bigg] \\
&- (NT)^{-2} \left(\sum_{i,j,k}^{N} \sum_{s,t,u}^{T} \mathrm{E}\left[\hat{r}_{it;q} \left(a_{it}^{*\prime} \tilde{\Sigma}_\xi a_{it}^*\right)^{-2} w_{it}^{(\xi)\prime}\right.\right. \\
&\times \left[W^{(\xi)\prime} W^{(\xi)}\right]^{-1} w_{js}^{(\xi)} \left(\eta_{js}^2 - \mathrm{E}\left[\eta_{js}^2\right]\right) \\
&\times \left.\left.\left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right) w_{ku}^{(\xi)\prime} \left[W^{(\xi)\prime} W^{(\xi)}\right]^{-1} w_{it}^{(\xi)} \eta_{it}\right]\right) \\
&\times \left(\sum_{i,k}^{N} \sum_{t,u}^{T} \mathrm{E}\left[\hat{r}_{it;q} \left(a_{it}^{*\prime} \tilde{\Sigma}_\xi a_{it}^*\right)^{-2} w_{it}^{(\xi)\prime}\right.\right. \\
&\times \left.\left[W^{(\xi)\prime} W^{(\xi)}\right]^{-1} w_{ku}^{(\xi)} \left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right) \eta_{it}\right]\Bigg) \\
= \;& O\left(N^{-3} + T^{-3}\right),
\end{aligned}
$$

because

$$
\begin{aligned}
&\sum_{i,j,k,l,m}^{N} \sum_{s,t,u,v,x}^{T} \mathrm{E}\left[\left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right)\left(\eta_{lv}^2 - \mathrm{E}\left[\eta_{lv}^2\right]\right)\eta_{it}\eta_{js}\left(\eta_{mx}^2 - \mathrm{E}\left[\eta_{mx}^2\right]\right)\right] \\
&- \left(\sum_{i,j,k}^{N} \sum_{s,t,u}^{T} \mathrm{E}\left[\left(\eta_{js}^2 - \mathrm{E}\left[\eta_{js}^2\right]\right)\left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right)\eta_{it}\right]\right)\left(\sum_{i,k}^{N} \sum_{t,u}^{T} \mathrm{E}\left[\left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right)\eta_{it}\right]\right) \\
= \;& O\left(N^2T^5 + N^5T^2\right).
\end{aligned}
$$

Similarly, the following equalities can be shown: $\mathrm{Var}\left[T_j\right] = O\left(N^{-j} + T^{-j}\right), \forall j \geq 4$, $\mathrm{Cov}\left[T_i, T_j\right] = O_p\left(N^{-\lceil (i+j)/2 \rceil} + T^{-\lceil (i+j)/2 \rceil}\right), \forall i, j \geq 4$, where $\lceil x \rceil$ is the smallest integer greater than or equal to x.

It also holds that $(NT)^{-1} \hat{R}' \hat{D}^{-1} A^{*(B)} \mathrm{Diag}_{i=1}^{N}\left[\hat{E}_i^{-1}\right] A^{*(B)\prime} \hat{D}^{-1} H$ has as q^{th} element

$$
(NT)^{-1} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{t=1}^{T} \hat{r}_{it;q} \left(a_{it}^{*\prime} \hat{\tilde{\Sigma}}_\xi a_{it}^*\right)^{-1} a_{it}^{*\prime} \left[\sum_{r=1}^{T} a_{ir}^* \left(a_{ir}^{*\prime} \hat{\tilde{\Sigma}}_\xi a_{ir}^*\right)^{-1} a_{ir}^{*\prime}\right]^{-1} a_{is}^* \left(a_{is}^{*\prime} \hat{\tilde{\Sigma}}_\xi a_{is}^*\right)^{-1} \eta_{is},
$$

$(NT)^{-1} \hat{R}' \hat{D}^{-1} A^{*(T)} \hat{G}_1^{-1} A^{*(T)\prime} \hat{D}^{-1} H$ has as q^{th} element

$$
(NT)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \hat{r}_{it;q} \left(a_{it}^{*\prime} \hat{\tilde{\Sigma}}_\xi a_{it}^*\right)^{-1} a_{it}^{*\prime} \left[\sum_{h=1}^{N} a_{ht}^* \left(a_{ht}^{*\prime} \hat{\tilde{\Sigma}}_\xi a_{ht}^*\right)^{-1} a_{ht}^{*\prime}\right]^{-1} a_{jt}^* \left(a_{jt}^{*\prime} \hat{\tilde{\Sigma}}_\xi a_{jt}^*\right)^{-1} \eta_{jt}
$$

and $(NT)^{-1}\hat{R}'\hat{D}^{-1}A^{*(T)}\hat{G}_1^{-1}\hat{H}_1\hat{G}_1^{-1}A^{*(T)\prime}\hat{D}^{-1}H$ has as q^{th} element

$$(NT)^{-1}\sum_{i,j}^{N}\sum_{s,t}^{T}\left\{\hat{r}_{it;q}\left(a_{it}^{*\prime}\hat{\tilde{\Sigma}}_{\xi}a_{it}^{*}\right)^{-1}a_{it}^{*\prime}\left[\sum_{h=1}^{N}a_{ht}^{*}\left(a_{ht}^{*\prime}\hat{\tilde{\Sigma}}_{\xi}a_{ht}^{*}\right)^{-1}a_{ht}^{*\prime}\right]^{-1}\right.$$
$$\left[\sum_{h=1}^{N}a_{ht}^{*}\left(a_{ht}^{*\prime}\hat{\tilde{\Sigma}}_{\xi}a_{ht}^{*}\right)^{-1}a_{ht}^{*\prime}\left[\sum_{r=1}^{T}a_{hr}^{*}\left(a_{hr}^{*\prime}\hat{\tilde{\Sigma}}_{\xi}a_{hr}^{*}\right)^{-1}a_{hr}^{*\prime}\right]^{-1}a_{hs}^{*}\left(a_{hs}^{*\prime}\hat{\tilde{\Sigma}}_{\xi}a_{hs}^{*}\right)^{-1}a_{hs}^{*\prime}\right]$$
$$\left.\left[\sum_{h=1}^{N}a_{hs}^{*}\left(a_{hs}^{*\prime}\hat{\tilde{\Sigma}}_{\xi}a_{hs}^{*}\right)^{-1}a_{hs}^{*\prime}\right]^{-1}a_{js}^{*}\left(a_{js}^{*\prime}\hat{\tilde{\Sigma}}_{\xi}a_{js}^{*}\right)^{-1}\eta_{js}\right)$$

From which the remark below Lemma D.1 can be verified, i.e. the terms of $\hat{R}'\Phi_J Y$, $J = B, C, D$, converge at the same rates as the corresponding terms of $\hat{R}'\hat{D}^{-1}Y$.

The variance of $(NT)^{-1}\hat{R}'\hat{D}^{-1}H$ is now equal to $O\left((NT)^{-1}+N^{-2}+T^{-2}\right)$ and $(NT)^{-1}\hat{R}'\hat{D}^{-1}H = o_p(1)$, by the Chebyshev inequality. Suppose that $N \sim T^{\kappa}$, then, since $(NT)^{-1}\hat{R}'\hat{D}^{-1}R = O_p(1)$, it holds that

$$\mathrm{Var}\left[\hat{\varphi}_{F2SA}-\varphi\right] = \begin{cases} O\left(N^{-2}\right) \text{ if } \kappa<1 \\ O\left(N^{-1}T^{-1}\right), \text{ if } \kappa=1 \\ O\left(T^{-2}\right), \text{ if } \kappa>1 \end{cases},$$

which can be summarized as

$$\min(N,T)\left(\hat{\varphi}_{F2SA}-\varphi\right) = O_p(1),$$

which proves part 2 of the theorem.

Remark. Note that in the second case it doesn't hold that

$$\mathrm{plim}_{N,T\to\infty}\left[(NT)^{-1}\Psi_{\Omega}\hat{R}'\left(\hat{\Omega}_{\eta}^{-1}-\Omega_{\eta}^{-1}\right)H\right] = 0,$$

and thus the equivalence between Aitken and feasible Aitken estimator cannot be proved.

References

[1] Ahn, S.C., Schmidt, P. (1995), "Efficient Estimation of Models for Dynamic Panel Data", *Journal of Econometrics*, **68**, p.5-27.

[2] Ahn, S.C., Schmidt, P. (1997), "Efficient Estimation of Dynamic Panel Data Models: Alternative Assumptions and Simplified Estimation", *Journal of Econometrics*, **76**(1-2), p.309-321.

[3] Ahn, S.C., Schmidt, P. (1999), "Estimation of Linear Panel Data Models Using GMM", in Mátyás, L. (ed.), *Generalized Method of Moments Estimation*, Cambridge University Press, Cambridge, p.211-247.

[4] Amemiya, T., MaCurdy, T. (1986) "Instrumental-Variable Estimation of an Error-Components model", *Econometrica*, **54**, p.869-881.

[5] Anderson, T.W. (1970), "Estimation of Covariance Matrices Which are Linear Combinations or Whose Inverses are Linear Combinations of Given Matrices", in Bose, R.C. (ed.), *Essays in Probability and Statistics*, Chapel Hill, University of North Carolina Press, p.1-24.

[6] Arellano, M., Bond, S. (1991),"Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, **58**, p.277-297.

[7] Arellano, M., Bover, O. (1995),"Another Look at the Instrumental Variable Estimation of Error-Components Models", *Journal of Econometrics*, **68**, p.29-51.

[8] Baltagi, B. H., Khanti-Akom, S. (1990), "On Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators", *Journal of Applied Econometrics*, **5**(4), p.401-406.

[9] Belzil, C., Hansen, J. (2002),"A Structural Analysis of the Correlated Random Coefficient Wage Regression Model", *CIRANO Série Scientifique*, **2002s-07**, Centre interuniversitaire de recherche en analyse des organisations, Montréal, pp.33.

[10] Breusch, T.S., Mizon, G.E., Schmidt, P. (1989), "Efficient Estimation Using Panel Data", *Econometrica*, **57**, p.695-700.

[11] Card, D. (2001), "Estimating the Return to Schooling: Progress on some Persistent Econometric Problems", *Econometrica*, **69**, p.1127-1160.

[12] Carroll, R.J., Ruppert, D. (1982), "Robust Estimation in Heteroskedastic Linear Models", *Annals of Statistics*, **10**(2), p.429-441.

[13] Cornwell, C., Rupert, P. (1988), "Efficient Estimation with Panel Data: an Empirical Comparison of Instrumental Variables Estimators", *Journal of Applied Econometrics*, **3**, p.149-155.

[14] Crockett, P.W. (1985), "Asymptotic Distribution of the Hildreth-Houck Estimator", *Journal of the American Statistical Association*, **80**, p.202-204.

[15] Fuller, W., Battese, G. (1973), "Transformations for Estimation of Linear Models with Error Structure", *Journal of the American Statistical Association*, **68**, p.626-632.

[16] Garen, J. (1984), "The Returns to Schooling: a Selectivity Bias Approach with a Continuous Choice Variable", *Econometrica*, **52**, p.1199-1218.

[17] Hausman, J.A., Taylor, W.E. (1981), "Panel Data and Unobservable Individual Effects", *Econometrica*, **49**, p.1377-1398.

[18] Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica*, **46**, p.931-959.

[19] Heckman, J., Vytlacil, E. (1998), "Instrumental Variables Methods for The Correlated Random Coefficient Model", *Journal of Human Resources*, **33**, p.974-987.

[20] Hildreth, C.,Houck, J.P. (1968), "Some Estimators for a Linear Model with Random Coefficients", *Journal of the American Statistical Association*, **63**, p.584-595.

[21] Hsiao, C. (1974), "Statistical Inference for a Model with Both Random Cross-Sectional and Time Effects", *International Economic Review*, **15**, p.12-30.

[22] Hsiao, C. (1975), "Some Estimation Methods for a Random Coefficient Model", *Econometrica*, **43**, p.305-325.

[23] Kelejian, H.H., Stephan, S.W. (1983), "Inference in Random Coefficient Panel Data Models: A Correction and Clarification of the Literature", *International Economic Review*, **24**(1), p.249-254.

[24] Maddala, G.S. (1971), "The Use of Variance Components Models in Pooling Cross-Section and Time Series Data", *Econometrica*, **39**, p.341-358.

[25] Mandy, D.M., Martins-Filho, C. (1994), "A Unified Approach to Asymptotic Equivalence of Aitken and Feasible Aitken Instrumental Variables Estimators", *International Economic Review*, **35**(4), p.957-979.

[26] Rao, C.R. (1965), "The Theory of Least Squares When the Parameters Are Stochastic and Its Application to the Analysis of Growth Curves", *Biometrika*, **52**, p.447-458.

[27] Searle, S.R., Casella, G., McCulloch, C.E. (1992), *Variance Components*, Wiley Series in Probability and Mathematical Statistics, Applied Probability and Statistics, Wiley, N.Y.

[28] Swamy, P.A.V.B. (1970), "Efficient Inference in a Random Coefficient Regression Model", *Econometrica*, **38**(2), p.311-323.

[29] White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.", *Econometrica*, **48**, p.817-838.

[30] Wooldridge, J.M. (1997), "On two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model", *Economics Letters*, **56**, p.129-133.

[31] Wooldridge, J.M. (2003a), "Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model", *Economics Letters*, **79**, p.129-133.

[32] Wooldridge, J.M. (2003b), "On the Robustness of Fixed Effects and Related Estimators in Correlated Random Coefficient Panel Data Models", *mimeo*.

[33] Wyhowski, D.J. (1994), "Estimation of a Panel-Data Model in the Presence of Correlation between Regressors and a 2-Way Error Component", *Econometric Theory*, **10**, p. 130-139.