Efficient ML and GMM estimation of panel data models with cross-sectional heteroskedasticity

Hugo Kruiniger*
Durham University

This incomplete version: 23 December 2014.

Abstract

This paper considers estimation of panel data models with (multiplicative) individual fixed effects in the variances of the errors, e.g., $\sigma_{i,t}^2 = \sigma_i^2$ or $\sigma_{i,t}^2 = \sigma_i^2 \lambda_t^2$. The cross-section dimension of the panel (N) is assumed to be large but the time dimension (T) can be small or large. The paper shows that under certain conditions, which depend on whether T is fixed or large, the common parameters in static (or stationary) and non-stationary dynamic linear panel data models can be consistently estimated by a *Weighted* First Difference Maximum Likelihood (FDML) estimator and a *Weighted* Random Effects or Fixed Effects ML (REML or FEML) estimator, respectively, and derives their asymptotic distributions. These estimators weigh the data with estimates of the σ_i^2 and are shown to be asymptotically efficient under joint N, T asymptotics and normality. We also discuss two-step Weighted Quasi ML estimators and Hybrid Quasi ML estimators that are generally still joint N, T consistent for the common parameters in non-stationary dynamic linear panel data models with arbitrary heteroskedasticity, i.e., when $\sigma_{i,t}^2 \neq \sigma_i^2 \lambda_t^2$. The paper then introduces *Individually Weighted* GMM (IWGMM) estimators that generalize the Minimum Distance estimator of Chamberlain (1982). Under normality and when T is fixed, the optimally weighted IWGMM estimators for variance parameters are more efficient than the corresponding Weighted ML estimators, whereas their unweighted ML and GMM counterparts are both efficient under (cross-sectional) homoskedasticity. Finally, Monte Carlo results show that the weighted estimators are more efficient than their unweighted counterparts when T is not too small and there is a significant degree of heteroskedasticity in the cross-section dimension of the panel.

JEL classification: C12, C13, C23.

Keywords: cross-sectional heteroskedasticity, dynamic panel data model, GMM, incidental parameters, Minimum Distance, Quasi Maximum Likelihood.

*Address: hugo.kruiniger@durham.ac.uk; Dept. of Economics, 23-26 Old Elvet, Durham DH1 3HY.

1 Introduction

This paper studies efficient estimation of common parameters in panel data models with (multiplicative) individual fixed effects in the variances of the errors, e.g. panel models with $\sigma_{i,t}^2 \equiv E(\varepsilon_{i,t}^2) = \sigma_i^2$ or $E(\varepsilon_{i,t}^2|z_{i,t}) = \sigma_i^2\lambda_t^2 \exp(z_{i,t}'\vartheta)$. We consider estimation of both 'static' models, including covariance stationary models, and 'non-stationary' dynamic panel models, i.e., models with arbitrary initial conditions, including a panel AR(1)-GARCH(1,1) model with $E(\varepsilon_{i,t}^2|\Im_{t-1}) = h_{i,t} = \sigma_i^2\widetilde{h}_{i,t}$ and $\widetilde{h}_{i,t} = \alpha_0 + \alpha_1(\varepsilon_{i,t-1}^2/\sigma_i^2) + \beta_1\widetilde{h}_{i,t-1}$, where the set $\Im_{t-1}$ contains information until time $t-1$. The cross-section dimension of the panel (N) is assumed to be large but the time dimension (T) can be small or large.

The paper first shows that after a first-difference transformation and under certain conditions the common parameters in static or covariance stationary linear panel data models with additive fixed individual effects in the mean and multiplicative individual fixed effects in the variance can be consistently estimated by Maximum Likelihood. We call the resulting estimator the Weighted First Difference Maximum Likelihood (FDML) estimator because it weights the data by estimates of the σ_i^2. The paper then shows that dynamic linear panel data models with arbitrary initial conditions, random effects (RE) or fixed effects (FE) in the mean and variances modelled as $\sigma_{i,t}^2 = \sigma_i^2\lambda_t^2$ can also be consistently estimated by (Quasi) Maximum Likelihood. We refer to these estimators as the Weighted (Quasi) REML estimator and the Weighted (Quasi) FEML estimator, respectively. In each case the conditions for consistency of the Weighted MLE are non-trivial when T is fixed and only $N \to \infty$. The paper also derives the large N, fixed T and the joint N, T asymptotic distributions of the three Weighted ML estimators. The estimators are shown to be asymptotically efficient under joint N, T asymptotics and normality. We also discuss two-step weighted Quasi ML estimators and Hybrid Quasi ML estimators that are generally still joint N, T consistent for the common parameters in non-stationary dynamic linear panel data models with arbitrary heteroskedasticity, i.e., when $\sigma_{i,t}^2 \neq \sigma_i^2\lambda_t^2$.

Next the paper introduces Individually Weighted GMM (IWGMM) estimators that generalize the Minimum Distance and GMM estimators of Chamberlain (1982) and Hansen (1982), respectively.[1] Under normality and when T is fixed, the optimally

[1] In addition to weighing the individual data, the IWGMM estimators also use a weight

weighted IWGMM estimators for the variance parameters in both static/stationary and nonstationary linear panel data models are asymptotically more efficient than the corresponding Weighted ML estimators, whereas their unweighted FDML and GMM counterparts are both efficient under cross-sectional (CS) homoskedasticity. Furthermore, under non-normality the optimally weighted IWGMM estimators for common parameters in these models are generally more efficient than the corresponding Weighted ML estimators.

The common parameters in the panel models can still be consistently estimated by estimators that ignore cross-section heteroskedasticity, i.e., by the unweighted ML and GMM estimators. However, Monte Carlo results confirm that the (optimal versions of the individually) weighted estimators are more efficient than their unweighted counterparts when T is not too small and there is a significant degree of heteroskedasticity in the cross-section dimension of the panel.

It has been known since Neyman and Scott (1948) that ML estimators of common parameters can be inconsistent or asymptotically inefficient in the presence of so-called incidental parameters. Kiefer (1980) discussed ML estimation of the panel regression model with fixed effects in the mean and homogenous unrestricted covariance matrices. He first showed that the standard ML estimator for the covariance matrix is inconsistent and then proposed a Conditional ML estimator for the slope parameters. MaCurdy (1982) has suggested to base the ML estimators for the common parameters on the likelihood function of the first-differences of the data. Chamberlain (1980) and Anderson and Hsiao (1982) introduced the REML estimator for the dynamic panel AR(1) model with arbitrary initial conditions. Hsiao et al. (2002) and Kruiniger (2001) introduced versions of the FEML estimator for this model. Kruiniger (2008) establishes consistency of the FDML estimator for the covariance stationary AR(1)/unit root panel model with fixed effects in the mean under CS homoskedasticity and derives its limiting distribution under various asymptotic plans for N and T. Alvarez and Arellano (2003) discusses the joint N, T asymptotic properties of the REMLE under homoskedasticity. Alvarez and Arellano (2004) and Kruiniger (2013) generalize the REMLE by allowing for time series heteroskedasticity; the latter also shows that the RE- and FEMLE are still large N, fixed T consistent for the common parameters under arbitrary heteroskedasticity and states the large N, fixed T asymptotic distributions of these Quasi MLEs, also for the

matrix to weigh the moment conditions just like any other GMM estimator.

unit root case. Hawakaya and Pesaran (2012) derive the asymptotic distributions of the unweighted Quasi FEMLE when $\sigma_{i,t}^2 = \sigma_i^2$. Finally, Bai (2013) obtains the REQMLE for the model with time series heteroskedasticity using a factor analytical approach and derives its joint N, T asymptotic properties. He notes that this unweighted REQMLE is still joint N, T consistent for the common parameters under arbitrary heteroskedasticity and can be extended to the case with time dummies. He also shows that the REQMLE is both large N, fixed T and joint N, T asymptotically efficient if the individual vectors of errors are i.i.d., the errors are also independent across time, and are either normal or, in case T is fixed, have finite $2 + \varsigma$ moments for some $\varsigma > 0$.

The literature has studied asymptotically efficient estimation of regression models for cross-sectional or time series data with heteroskedasticity of unknown form. Cragg (1983) proposed an instrumental variable estimator for the slope parameters of a linear regression model that is at least as efficient asymptotically as the OLS estimator. Robinson (1987) discusses an adaptive estimation method that is based on nearest neighbour nonparametric regression. In contrast to Cragg's estimator and our approach, the adaptive estimation method assumes that the variances depend on (some of) the regressors. Li and Stengos (1994) discuss adaptive estimators for RE and FE panel data models with heteroskedasticity of unknown form.

Meghir and Windmeijer (1999) discuss a GMM estimator for the autoregressive parameter in a panel AR(1) model with multiplicative individual fixed effects in the variances of the errors. However, their estimator does not weigh the data with estimates of the σ_i^2 and is therefore not asymptotically efficient under CS heteroskedasticity. Pakel et al. (2011) consider a m-profile composite likelihood estimator for a panel GARCH model with individual specific unconditional variances. We will consider standard ML estimation of the same panel GARCH model.

The plan of the paper is as follows. Section 2 discusses ML estimation of static panel data models with CS heteroskedasticity. Section 3 considers ML estimation under general heteroskedasticity. Section 4 discusses ML estimation of nonstationary dynamic panel data models with CS heteroskedasticity. Section 5 considers GMM estimation under multiplicative CS heteroskedasticity. Section 6 reports the results of various Monte Carlo experiments while section 7 concludes. The appendices contain the proofs of the theorems.

2 ML estimation of static panel data models with cross-sectional heteroskedasticity

In this section we will consider ML estimation of the following panel regression model with fixed effects in both the mean and the variance:

$$
\begin{aligned}
& y_i - X_i\beta - \mu_i\iota = \varepsilon_i, && (1)\\
\sigma_i^{-1}(V(W_i,\theta_0))^{-1/2}\varepsilon_i, \quad & i = 1, ..., N, \text{ are } i.i.d.(0, I), &&\\
W_i, \quad & i = 1, ..., N, \text{ are } i.i.d., \text{ and} &&\\
\sigma_i^{-1}(V(W_i,\theta_0))^{-1/2}\varepsilon_i \perp W_i, \quad & i = 1, ..., N, &&
\end{aligned}
$$

where the T-vector y_i and the $(T\times K)$ matrix X_i contain the T observations for individual i on the dependent variable and K regressors, ε_i is a T-vector of errors and ι is a T-vector of ones; β contains the true common (i.e., common) slope parameters; θ_0 contains the true common covariance parameters; and μ_i and σ_i^2, $i = 1, ..., N$, are true incidental mean and variance parameters, respectively. Note that the individual variance effect σ_i^2 enters the formula of the covariance matrix in a multiplicative way. This parametrization of the covariance matrix is natural given the interpretation of the variance as a scaling parameter. We will assume that the elements of $V(W_i,\theta) = V_i(\theta) = V_i$ are smooth functions of the elements of θ. The framework in (1) includes generalizations of the covariance stationary autoregressive panel data model discussed in Kruiniger (2008). For instance, V_i could be the covariance matrix of heteroskedastic AR(1) errors, i.e., $V_{i,s,t}(\theta_0) = \widetilde{\lambda}^2_{\min(s,t)}\rho^{|t-s|}$, $s, t = 1, ..., T$, with $|\rho| < 1$, $\widetilde{\lambda}^2_t = \rho^2\widetilde{\lambda}^2_{t-1} + \lambda^2_t$, $t = 2, ..., T$, a normalisation e.g. $\lambda_T = 1$, and $\theta_0 = (\widetilde{\lambda}^2_1\ \lambda^2_2\ \lambda^2_3\ ...\ \lambda^2_{T-1}\ \rho)'$.

To focus the discussion we will assume that $\beta = 0$. After first-differencing the model to remove heterogeneity in the mean, we obtain

$$\sigma_i^{-1}(\Phi(W_i,\theta_0))^{-1/2}Dy_i,\ i = 1, ..., N, \text{ are } i.i.d.(0, I), \tag{2}$$

where $\Phi(W_i,\theta) = \Phi_i(\theta) = \Phi_i = DV_iD'$ and D denotes a $(T-1\times T)$ matrix with $D_{k,k} = -1$ and $D_{k,k+1} = 1$, $k = 1, ..., T-1$, and $D_{k,l} = 0$ elsewhere. Note that $(\Phi(\theta_0))^{-1/2}Dy_i \perp W_i$.

The conditional Gaussian log-likelihood function of the Dy_i given the W_i reads

$$\sum_{i=1}^{N} l(Dy_i;\theta, s_i^2|W_i) = -\frac{N(T-1)}{2}\log(2\pi) - \frac{(T-1)}{2}\sum_{i=1}^{N}\log s_i^2 \tag{3}$$
$$-\frac{1}{2}\sum_{i=1}^{N}\log|DV_iD'| - \sum_{i=1}^{N}\frac{1}{2s_i^2}(y_i'D'\left(DV_iD'\right)^{-1}Dy_i).$$

Let $Z_i = Dy_iy_i'D'$. Then we can write the likelihood equations as

$$\frac{1}{2}\sum_{i=1}^{N} tr(\frac{d\Phi_i^{-1}}{d\theta_l}(\Phi_i - \frac{Z_i}{s_i^2})) = 0,\ \ l = 1, ..., \dim(\theta), \tag{4}$$

and

$$\frac{T-1}{2s_i^2} - \frac{1}{2s_i^4}tr(\Phi_i^{-1}Z_i) = 0,\ \ i = 1, ..., N. \tag{5}$$

Solving the N equations in (5) for s_i^2, $i = 1, ..., N$, yields

$$\widetilde{\sigma}_i^2(\theta) = \frac{1}{T-1}tr([\Phi_i(\theta)]^{-1}Z_i),\ \ i = 1, ..., N. \tag{6}$$

After substituting the $\widetilde{\sigma}_i^2(\theta)$ for the s_i^2 in (4), we obtain the following system of p concentrated likelihood equations:

$$-\frac{1}{2}N^{-1}\sum_{i=1}^{N}[\frac{dvec\Phi_i}{d\theta'}]'(\Phi_i^{-1}\otimes\Phi_i^{-1})vec(\Phi_i - \frac{Z_i}{\widetilde{\sigma}_i^2(\theta)}) = 0. \tag{7}$$

These equations define the First Difference ML Estimator (FDMLE) for θ_0, viz. $\widehat{\theta}_{FDMLE}$. Note that when $T = 2$, $\Phi_i(\theta) - Z_i/\widetilde{\sigma}_i^2(\theta) = 0$ for any θ, so we need at least that $T \geq 3$.

We need to introduce some additional notation. Let Θ be the parameter space for θ and let $\|A\| = tr(A'A)^{1/2}$ be the Euclidean norm of a matrix A. Moreover, let $S_i(\theta) = \Phi_i^{-1/2}Z_i\Phi_i^{-1/2}/((T-1)\widetilde{\sigma}_i^2(\theta))$.

Note that $E(\Phi_i(\theta_0) - Z_i/\sigma_i^2) = 0$. If the elements of $(\Phi_i(\theta_0))^{-1/2}Dy_i$ are symmetrically i.i.d. for each $i \in \{1, ..., N\}$, then it is easily verified that $E((T-1)S_i(\theta_0)) = I$ and $E(\Phi_i(\theta_0) - Z_i/\widetilde{\sigma}_i^2(\theta_0)) = 0$.[2]

In order to derive the asymptotic properties of the FDMLE we make the following assumption:

[2] If $Dy_i|W_i \sim N(0, \sigma_i^2\Phi_i(\theta_0))$, $i = 1, ..., N$, then the diagonal elements of $S_i(\theta_0)$ follow a Beta$(\frac{1}{2}, \frac{T-2}{2})$ distribution and the non-diagonal elements of $S_i(\theta_0)$ are symmetrically distributed around zero.

Assumption A: Let $g_i(\theta) = -(\frac{dvec\Phi_i}{d\theta'})'([\Phi_i(\theta)]^{-1} \otimes [\Phi_i(\theta)]^{-1})vec(\Phi_i(\theta) - Z_i/\widetilde{\sigma}_i^2(\theta))$.

(i) $\theta_0 \in int(\Theta)$, $\dim(\theta)$ is fixed, and Θ is compact.

(ii) $\Phi_i(W_i, \theta)$ is a PDS matrix $\forall \theta \in \Theta$ and $\forall vec(W_i) \in support(vec(W_i))$.

(iii) $(((\sigma_i^2\Phi_i(\theta_0))^{-1/2}Dy_i)'\ vec(W_i)')'$, $i = 1, \ldots, N$, are i.i.d.

(iv) $(\Phi_i(\theta_0))^{-1/2}Dy_i \perp W_i\ \forall i \in \{1, ..., N\}$.

(v) $g_i(\theta)$ is a continuous function at each $\theta \in \Theta$ uniform in T a.s.

(vi) $E(g_i(\theta)) = 0$ iff $\theta = \theta_0$ uniform in T.

(vii) $\sup_{\theta\in\Theta} \left\| (NT)^{-1} \sum_{i=1}^N g_i(\theta) - T^{-1}E(g_i(\theta)) \right\| \xrightarrow{p} 0$.

(viii) If T is fixed: $[(\sigma_i^2\Phi(\theta_0))^{-1/2}Dy_i]_t$ are symmetrically i.i.d. $\forall i, t$.

We have the following result:

Theorem 1 *Suppose that assumption A holds. Then the FDMLE for θ in model (1) is consistent when $N \to \infty$ irrespective of whether T is fixed or $T \to \infty$.*

The result follows by applying theorems 2.1 and 2.6 in Newey and McFadden (1994), henceforth NMcF. When T is fixed we can replace A(vii) by $E\left(\sup_{\theta\in\Theta} \|g_i(\theta)\|\right) < \infty$. When $T \to \infty$, A(vii) may be verified by using a uniform in θ version of results in Phillips and Moon (1999), e.g. their Theorem 1 or Corollary 1.

Thus the FDMLE is consistent for the common covariance parameters under unrestricted multiplicative CS heteroskedasticity. Kiefer and Wolfowitz (1956) have considered ML estimation of a similar model where the σ_i^2 are i.i.d.

Let $F_i(\theta) = [\frac{dvec\Phi_i}{d\theta'}]'(\Phi_i^{-1/2} \otimes \Phi_i^{-1/2})$. Since $vec(\Phi_i(\theta) - Z_i/\widetilde{\sigma}_i^2(\theta)) = (\Phi_i^{1/2} \otimes \Phi_i^{1/2})vec[I - (T-1)S_i(\theta)]$, it is easily seen that $g_i(\theta) = -F_i(\theta)vec[I - (T-1)S_i(\theta)]$. Furthermore, let $G(\theta_0) = (T-1)^2T^{-1}E\{F_i(\theta_0)Cov(vec(S_i(\theta_0)))F_i'(\theta_0)\}$. To establish asymptotic normality of the FDMLE we add the following assumption:[3]

Assumption B:

(i) $g_i(\theta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of θ_0 uniform in T a.s.

(ii) $(NT)^{-1/2} \sum_{i=1}^N g_i(\theta_0) \xrightarrow{d} N(0, G(\theta_0))$.

(iii) $\sup_{\theta\in\mathcal{N}} \left\| (NT)^{-1} \sum_{i=1}^N \frac{dg_i(\theta)}{d\theta'} - H(\theta) \right\| \xrightarrow{p} 0$ where $H(\theta) = E(T^{-1}\frac{dg_i(\theta)}{d\theta'})$.

(iv) $H(\theta_0)$ is nonsingular.

[3]Below, if $T \to \infty$, then replace $G(\theta_0)$ and $H(\theta)$ by $\lim_{T\to\infty} G(\theta_0)$ and $\lim_{T\to\infty} H(\theta)$.

Theorem 2 *Suppose that assumptions A and B hold. Then the large N, fixed T asymptotic distribution of the FDMLE for θ in (1) is given by*

$$\sqrt{N}(\widehat{\theta}_{FDMLE} - \theta_0) \xrightarrow{d} N(0, T^{-1}H(\theta_0)^{-1}G(\theta_0)H(\theta_0)^{-1}). \tag{8}$$

The result follows from theorem 3.2 in NMcF. $\widehat{\theta}_{FDMLE}$ is consistent by assumption A. Note that the $vecS_i(\theta_0)$ are i.i.d. with finite second moments because all its elements lie in $[-1, 1]$. Hence if Φ_i does not depend on W_i, then assumption B(ii) is clearly satisfied. $H(\theta)$ is continuous at θ_0 because of assumption B(i). In appendix A we show that $H(\theta_0) = -T^{-1}E\{F_i(\theta_0)[I - (T-1)(vecS_i(\theta_0)(vecS_i(\theta_0))')]F_i'(\theta_0)\}$.

Theorem 3 *Suppose that assumptions A and B hold and that the number of observations that are available to estimate each element of θ grows at rate T. Then the joint N, T asymptotic distributions of the FDMLE for θ and σ_i^2, $i = 1, \ldots, N$, in (1) are given by*

$$\sqrt{NT}(\widehat{\theta}_{FDMLE} - \theta_0) \xrightarrow{d} N(0, \lim_{T\to\infty}(H(\theta_0)^{-1}G(\theta_0)H(\theta_0)^{-1})) \text{ and}$$
$$\sqrt{T}(\widehat{\sigma}^2_{i,FDMLE} - \sigma_i^2) \xrightarrow{d} N(0, Asyvar(\widehat{\sigma}^2_{i,FDMLE})),\ i = 1, \ldots, N, \tag{9}$$

where $\lim_{T\to\infty} H(\theta_0) = \lim_{T\to\infty} -T^{-1}(\frac{dvec\Phi_i}{d\theta'}|_{\theta_0})'([\Phi_i(\theta)]^{-1} \otimes [\Phi_i(\theta)]^{-1})(\frac{dvec\Phi_i}{d\theta'}|_{\theta_0})$. *Furthermore, if* $\sigma_i^{-1}(\Phi(W_i, \theta_0))^{-1/2}Dy_i \sim N(0, I)$ *for* $i = 1, \ldots, N$, *then* $\lim_{T\to\infty} G(\theta_0) = \lim_{T\to\infty} -2H(\theta_0)$, $\widehat{\theta}_{FDMLE}$ *is asymptotically efficient and* $Asyvar(\widehat{\sigma}^2_{i,FDMLE}) = 2\sigma_i^4$.

The first result follows again from theorem 3.2 in NMcF. Note that $\text{plim}_{T\to\infty}\widehat{\sigma}^2_{i,FDMLE} = \sigma_i^2$. Note also that theorem 3 rules out period-specific parameters in $V_i(\theta)$. Relaxing this assumption is possible but affects the rate of convergence of the FDMLE for the period-specific parameters. For instance, consider the following example: $\sigma_i^2 V_i(\theta_0) = \sigma_i^2 diag(\lambda_1^2, \lambda_2^2, ..., \lambda_{T-1}^2, 1)$. Then $\sqrt{T}(\widehat{\sigma}^2_{i,FDMLE} - \sigma_i^2) \xrightarrow{d} N(0,\ Asyvar(\widehat{\sigma}^2_{i,FDMLE}))$, $i = 1, \ldots, N$, and $\sqrt{N}(\widehat{\lambda}^2_{t,FDMLE} - \lambda_t^2) \xrightarrow{d} N(0, Asyvar(\widehat{\lambda}^2_{t,FDMLE}))$, $t = 1, \ldots, T-1$, and if $\varepsilon_i \sim N(0, \sigma_i^2 V_i(\theta_0))$ for $i = 1, \ldots, N$, $Asyvar(\widehat{\lambda}^2_{t,FDMLE}) = 2\lambda_t^4$. If we estimate a two-way heteroskedastic panel AR(1) model, so that $s_i^2 V_{i,k_1,k_2}(\theta) = s_i^2 l^2_{\min(k_1,k_2)} r^{|k_1-k_2|}$, $k_1, k_2 = 1, ..., T$, with $l_T = 1$ and $|\rho| < 1$, then under time series homoskedasticity we have $\sqrt{NT}(\widehat{\rho}_{FDMLE} - \rho) \xrightarrow{d} N(0, 1-\rho^2)$ just as in the fully homoskedastic case (cf. Kruiniger, 2008, and Bai, 2013).

We conjecture that if the values of σ_i^2, $1, ..., N$, were known, $\sigma_i^{-1}(V_i(\theta_0))^{-1/2}\varepsilon_i \sim N(0, I)$ and T were fixed, then the corresponding FDMLE would be an asymptotically efficient *Fixed* Effects estimator for θ_0, although it would not attain the generalized Cramér-Rao lowerbound of Neyman and Scott (1948) or the Fisher information bound (cf. Kruiniger, 2008)).

3 ML estimation under general heteroskedasticity

Consider the following simplified model with completely unrestricted variance parameters:

$$y_i = \varepsilon_i \sim N(0, \Gamma_i V \Gamma_i),\ i = 1, ..., N, \tag{10}$$

where $\Gamma_i = diag(\sigma_{i,1}, ..., \sigma_{i,T})$. For convenience, let $T = 2$. The results that we obtain for this model can be generalized to cases where $T > 2$. W.l.o.g. we parametrize V as

$$V = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{11}$$

Let us define

$$\Sigma_i \equiv \begin{bmatrix} \sigma_{i,1}^2 & \rho\sigma_{i,1}\sigma_{i,2} \\ \rho\sigma_{i,1}\sigma_{i,2} & \sigma_{i,2}^2 \end{bmatrix}, \text{ and } X_i \equiv \begin{bmatrix} y_{i,1}^2 & y_{i,1}y_{i,2} \\ y_{i,1}y_{i,2} & y_{i,2}^2 \end{bmatrix}. \tag{12}$$

The likelihood equations for the $\sigma_{i,t}$ in model (10), $tr(\frac{d\Sigma_i^{-1}}{d\sigma_{i,t}}(\Sigma_i - X_i)) = 0$, $i = 1, ..., N$, $t = 1, 2$ can be rewritten as $\rho^2(1 - \frac{X_{i,12}}{\rho\sigma_{i,1}\sigma_{i,2}}) = 1 - \frac{X_{i,tt}}{\sigma_{i,t}^2}$. It follows that $X_{i,11}/\sigma_{i,1}^2 = X_{i,22}/\sigma_{i,2}^2$ and

$$\widehat{\sigma}_{i,t}^2(\rho) \equiv (1 - \rho R_i)\frac{X_{i,tt}}{(1 - \rho^2)} = \sigma_{i,t}^2, \tag{13}$$

where $R_i \equiv X_{i,12}/\sqrt{X_{i,11}X_{i,22}}$. Using these results we can simplify the likelihood equation for ρ, $\sum_{i=1}^N tr(\frac{d\Sigma_i^{-1}}{d\rho} \times (\Sigma_i - X_i)) = 0$. We find that

$$\rho = N^{-1}\sum_{i=1}^N \left(X_{i,12}/\sqrt{\sigma_{i,1}^2\sigma_{i,2}^2}\right). \tag{14}$$

Replacing the $\sigma_{i,t}^2$ in (14) by $\widehat{\sigma}_{i,t}^2(\rho)$, we obtain an equation that defines the ML estimator

$\widehat{\rho}$ for ρ

$$\widehat{\rho} = N^{-1} \sum_{i=1}^{N} \left(R_i \frac{(1-\widehat{\rho}^2)}{(1-\widehat{\rho} R_i)} \right). \quad (15)$$

Since $R_i^2 = 1$, it is easily seen that $(1-\widehat{\rho}^2)/(1-\widehat{\rho}R_i) = 1+\widehat{\rho}R_i$. Condition (15) therefore implies that $N^{-1}\sum_{i=1}^{N} R_i = 0$ which in most cases is not true. We conclude that the ML estimator for the common correlation parameter ρ is no longer consistent if the variance parameters are left completely unrestricted. The inconsistency of $\widehat{\rho}$ is due to the incidental parameters $\sigma_{i,t}^2$. If in (14) we would replace $\sigma_{i,t}^2$ by $\widetilde{\sigma}_{i,t}^2(\rho) = (1-\rho(N^{-1}\sum_{i=1}^{N} R_i))\frac{X_{i,tt}}{(1-\rho^2)}$ and solve for ρ, we would obtain the solution $\widetilde{\rho} = N^{-1}\sum_{i=1}^{N} R_i$, which is a consistent estimator for ρ. Note that $\widetilde{\sigma}_{i,t}^2(\widetilde{\rho}) = X_{i,tt}$.

4 ML estimation of nonstationary dynamic panel data models with cross-sectional heteroskedasticity

4.1 The panel AR(1) model

In this section we will consider ML estimation of versions of the following panel AR(1) model:

$$y_{i,t} = \rho y_{i,t-1} + \eta_i + \varepsilon_{i,t}, \text{ where } \eta_i = (1-\rho)\mu_i, \quad (16)$$

for $i = 1, ..., N$ and $t = 1, ..., T$. We assume that $|\rho| < 1$.[4]

The vectors of idiosyncratic errors $\varepsilon_i = (\varepsilon_{i,1} \ ... \ \varepsilon_{i,T})'$ are independently distributed across individuals and satisfy the following Standard Assumptions, SAk, for $k = 2$ or $k = 4$:

$$E(\varepsilon_{i,t}) = 0 \text{ and } E\left|\varepsilon_{i,t}\right|^{k+\varsigma} < \infty \text{ for } i = 1, ..., N \text{ and } t = 1, ..., T, \quad (17)$$

where $\varsigma \geq 0$ is an arbitrarily small constant. In the sequel SA2 is denoted by SA.

The individual effects can often be treated as random effects. In this case we make the following Random Effects Assumptions, REAk, for $k = 2$ or $k = 4$:

$$\sigma_{y,i}^{-1}(y_{i,0} \ \eta_i)', \ i = 1, ..., N, \text{ are i.i.d. with } \sigma_{y,i}^2 = Var(y_{i,0}), \ E\left|y_{i,0}\right|^{k+\varsigma} < \infty, \quad (18)$$

$$E(\mu_i) = 0, \ \sigma_{\mu,i}^2 = E(\mu_i^2), \ E\left|\mu_i\right|^{k+\varsigma} < \infty, \text{ and } E(\mu_i y_{i,0}) = \sigma_{\mu y,i}. \quad (19)$$

[4]The results presented below can be extended to $|\rho| \leq 1$ but when $\rho = 1$, the estimators discussed below will converge at a slower rate, $N^{1/4}$, and have non-normal limiting distributions under time-series homoskedasticity (cf. Kruiniger, 2013). Note that the parametrization $\eta_i = (1-\rho)\mu_i$ rules out a discontinuity in the data generating process at $\rho = 1$.

In addition, we let $\sigma^2_{\eta,i} = E(\eta_i^2)$ and $\sigma_{\eta y,i} = E(\eta_i y_{i,0})$. The assumption $E\left|\mu_i\right|^{k+\varsigma} < \infty$ (for $k = 2$ or $k = 4$) ensures that under covariance stationarity the means of the data, i.e., $\eta_i/(1-\rho) = \mu_i$, $i = 1, ..., N$, are drawn from a distribution with a finite variance rather than a variance that tends to infinity when ρ approaches one.

Unlike the RE ML estimators, the FE ML estimators only exploit data in first differences. This reflects the fact that the FE approach entails making minimal assumptions about the μ_i and the $y_{i,0}$. In the FE case we assume that $v_{i,0} \equiv y_{i,0} - \mu_i$, $i = 1, ..., N$, satisfy a Fixed Effects Assumption, FEAk, for $k = 2$ or $k = 4$ (cf. Kruiniger, 2001):

$$\sigma^{-1}_{v0,i} v_{i,0}, \; i = 1, ..., N, \text{ are i.i.d. with } \sigma^2_{v0,i} = Var(v_{i,0}) \text{ and } E\left|v_{i,0}\right|^{k+\varsigma} < \infty. \tag{20}$$

The i.i.d. assumption in (20) is not in the spirit of FE and is only made for presentational convenience. It can be relaxed. Below we sometimes allow $\sigma^{-1}_{v0,i} v_{i,0}$, $i = 1, ..., N$ to be i.h.d.

Suppose that $y_{i,0}$ depends on μ_i in a linear fashion, i.e. $y_{i,0} = E(y_{i,0}) + \alpha_1 \mu_i + \varepsilon_{i,0}$ with $E(\varepsilon_{i,0}) = 0$, and $\mu_i \perp \varepsilon_{i,0}$. In the important case that $\alpha_1 = 1$, we have $\mu_i \perp v_{i,0}$ and FEAk does not impose any restrictions on μ_i and $y_{i,0}$ other than those on $y_{i,0} - \mu_i$. However when $\alpha_1 \neq 1$, FEAk implies restrictions on the μ_i themselves.

In the sequel REA2 and FEA2 are denoted by REA and FEA, respectively.

For both the RE and the FE versions of the model we assume that

$$\varepsilon_{i,s} \perp \varepsilon_{i,t} \text{ for } i = 1, ..., N \text{ and } t \neq s. \tag{21}$$

Furthermore, for the RE versions of the model we assume that

$$\varepsilon_{i,t} \perp y_{i,0} \text{ and } \varepsilon_{i,t} \perp \eta_i \text{ for } i = 1, ..., N \text{ and } t = 1, ..., T, \tag{22}$$

whereas in the FE case we assume that

$$v_{i,0} \perp \varepsilon_{i,t} \text{ for } i = 1, ..., N \text{ and } t = 1, ..., T. \tag{23}$$

Finally, most of our discussion below is related to versions of the model that allow for heteroskedasticity of the $\varepsilon_{i,t}$ in both dimensions that satisfies the following equation:

$$\sigma^2_{i,t} = E(\varepsilon^2_{i,t}) = \sigma^2_i \lambda^2_t < \infty, \text{ for } i = 1, ..., N \text{ and } t = 1, ..., T, \text{ with } \lambda_1 = 1. \tag{24}$$

In some cases we make the stronger assumption of Time Series Homoskedasticity (TSH):

$$E(\varepsilon_{i,t}^2) = \sigma_i^2 < \infty, \text{ for } i = 1, ..., N \text{ and } t = 1, ..., T. \tag{25}$$

We may also assume Constant Variance Ratios (CVR), i.e., in the RE versions we may assume

$$\sigma_{y,i}^2 \sigma_i^{-2} = \sigma_y^2, \ \sigma_{\eta,i}^2 \sigma_i^{-2} = \sigma_\eta^2 \text{ and } \sigma_{\eta y,i} \sigma_i^{-2} = \sigma_{\eta y}, \text{ for } i = 1, ..., N, \tag{26}$$

while in the FE versions of the model we may assume

$$\sigma_{v_{0,i}}^2 \sigma_i^{-2} = \sigma_{v_0}^2, \text{ for } i = 1, ..., N, \tag{27}$$

so that the models have only N incidental variance parameters σ_i^2, $i = 1, ..., N$.

4.2 ML estimators

Direct application of the Maximum Likelihood method to the nonstationary panel AR(1) model with RE will generally yield an inconsistent estimator for ρ due to correlation between the individual effects (η_i) and the regressors ($y_{i,t-1}$, $t = 1, ..., T$). However, a consistent ML estimator for ρ can be obtained after reformulating the model. Following Chamberlain (1980) we can decompose the η_i into a term that depends on the initial observation, $y_{i,0}$, and a term that does not: [5]

$$\eta_i = \pi(1-\rho) y_{i,0} + (1-\rho) v_i, \ i = 1, ..., N, \tag{28}$$

where v_i is a new individual effect with $E(v_i) = 0$, $\pi(1-\rho) = \text{plim}_{N\to\infty} \sum_{i=1}^N (\sigma_{y,i}^{-2} \eta_i y_{i,0} / \sum_{i=1}^N (\sigma_{y,i}^{-2} y_{i,0}))$ and $\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N (\sigma_{y,i}^{-2} y_{i,0} v_i) = 0$.

Let $y_i = (y_{i,1} \ ... \ y_{i,T})'$ and $y_{i,-1} = (y_{i,0} \ ... \ y_{i,T-1})'$ and let ι denote a vector of ones. Then using the decomposition of the 'correlated effects' η_i given in (28), we can rewrite the panel AR(1) model with RE as

$$\begin{aligned} y_i &= \rho y_{i,-1} + \pi(1-\rho) y_{i,0} \iota + u_i, \quad \text{where} \\ u_i &= (1-\rho) v_i \iota + \varepsilon_i \quad \text{with} \quad E(\varepsilon_i \varepsilon_i') = \sigma_i^2 \Psi(\zeta) = \sigma_i^2 diag(1, \lambda_2^2, \lambda_3^2, ..., \lambda_T^2). \end{aligned} \tag{29}$$

[5] For the sake of a simple exposition we assume that $E(y_{i,0}) = 0$. A situation where $E(y_{i,0}) \neq 0$ can be handled by including an intercept term in (28) and in (29).

Let $\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}(v_i^2/\sigma_i^2)=\sigma_v^2$ and $\Phi=(1-\rho)^2\sigma_v^2\iota\iota'+\Psi(\zeta)$. Then it is easily verified that $E(y_{i,0}\iota'\Phi^{-1}u_i)=0$ and $E(y_{i,-1}'\Phi^{-1}u_i)=0$, cf. Blundell and Bond (1998). Note that when assumption CVR, i.e., (26) holds, then $E(v_i^2/\sigma_i^2)=\sigma_v^2$ for all $i\in\{1,...,N\}$. After adding the assumption that the scaled error components are i.i.d. and normal, i.e., $\sigma_i^{-1}u_i\sim i.i.d.\ N(0,\Phi)$, application of the ML method to (29) yields the (Weighted) RE ML Estimator for ρ, π, σ_v^2, $\zeta=(\lambda_2^2\ ...\ \lambda_T^2)'$ and σ_i^2, $i=1,...,N$.

When calculating the REMLE it is convenient to use the reparameterization $\tilde{\pi}=\pi(1-\rho)$ and $\widetilde{\sigma}_v^2=(1-\rho)^2\sigma_v^2$. Let $\theta_0=(\rho,\tilde{\pi},\widetilde{\sigma}_v^2,\zeta')'$. Then the log-likelihood function for the above model will be denoted by $l_{RE}(\theta,s_1^2,...,s_N^2)=\sum_{i=1}^{N}l_{RE,i}(\theta,s_i^2)$ where $\theta=(r,\widetilde{p},\widetilde{s}_v^2,z')'$ and $l_{RE,i}(\theta,s_i^2)=l_{RE,i}(y_i^+;\theta,s_i^2)$ is the contribution to $l_{RE}(\theta,s_1^2,...,s_N^2)$ from 'individual' i. The REMLE for θ_0 will be denoted by $\widehat{\theta}_{RE}$ or simply by $\widehat{\theta}$.

Both Hsiao et al. (2002) and Kruiniger (2001) have derived the FEMLE, respectively, for the model with $\sigma_i^2=\sigma^2$, $i=1,...,N$. Following the exposition in Kruiniger (2001), one can obtain the (Weighted) FE MLE for the model with CS heteroskedasticity by replacing μ_i in (16) by $y_{i,0}+v_i$, and imposing that the $\sigma_i^{-1}(v_i\ \varepsilon_i')$ are i.i.d. and normal with $v_i\perp\varepsilon_i$.[6] This amounts to imposing the restriction $\pi=1$ on the model in (29) and leads to the following formulation of the nonstationary panel AR(1) model with FE:

$$\begin{aligned} y_i &= \rho y_{i,-1}+(1-\rho)y_{i,0}\iota+u_i, \quad \text{where} \\ u_i &= (1-\rho)v_i\iota+\varepsilon_i \quad \text{with} \quad E(\varepsilon_i\varepsilon_i')=\sigma_i^2\Psi(\zeta), \end{aligned} \tag{30}$$

and where $v_i=-v_{i,0}$ satisfy assumption (23). After imposing $\sigma_i^{-1}u_i\sim i.i.d.\ N(0,\Phi)$ with $\Phi=\widetilde{\sigma}_v^2\iota\iota'+\Psi(\zeta)$, application of the ML method to (30) yields the (Weighted) FE ML Estimator for ρ, $\widetilde{\sigma}_v^2$, $\zeta=(\lambda_2^2\ ...\ \lambda_T^2)'$ and σ_i^2, $i=1,...,N$. The log-likelihood function for the above model will be denoted by $l_{FE}(\theta,s_1^2,...,s_N^2)$ where in this case $\theta=(r,\widetilde{s}_v^2,z')'$.

Let $\widetilde{\sigma}_i^2(\theta)=T^{-1}tr([\Phi(\theta)]^{-1}u_i(\theta)u_i'(\theta))$ with $u_i(\theta)=y_i-ry_{i,-1}-\widetilde{p}y_{i,0}\iota$ and let $\widetilde{\sigma}_i^2=\widetilde{\sigma}_i^2(\theta_0)$, $i=1,...,N$. Also let $\Theta\subset(-1,1)\times\mathbb{R}^{(\dim(\theta)-T-1)}\times[a,b]^T$ be a compact set with $b>a>0$ and let $\sigma_i^2\in[a,b]\ \forall i\in\{1,...,N\}$. To derive the asymptotic properties of the RE and FE (Quasi) MLEs when $\sigma_i^{-1}u_i$, $i=1,...,N$, are i.i.d., we may use

Assumption C:

[6] For the sake of a simple exposition we assume that $E(v_i)=0$. A situation where $E(v_i)\neq 0$ can be handled by including an intercept term in (30).

(i) $\sigma_i^{-1}\varepsilon_i$, $i = 1, ..., N$, are i.i.d., SA and REA or FEA with $\varsigma = 0$, $E(v_i) = 0$, (21), (22) or (23), (24) and (26) or (27) (ii) $v_i \perp y_{i,0}$ (iii) if T is fixed: (a) $[(\sigma_i^2\Phi(\theta_0))^{-1/2}u_i]_t$ are symmetrically i.i.d. $\forall i, t$ and (b) $E(\sup_{\theta\in\Theta}(\sigma_i/\widetilde{\sigma}_i(\theta))) < \infty$ (iv) if $T \to \infty$: SA4 with $\varsigma = 0$.

Note that assumption C(i) imposes CVR but allows the u_i, $i = 1, ..., N$, to be non-Gaussian. Assumptions C(ii) and C(iii-b) (and C$'$(ii) and D(b) below) are not needed for the FE(Q)MLE. Assumption C(iii) involves a strong assumption about the distribution of the u_i. However, it is satisfied when, for instance, $u_i \sim N(0, \sigma_i^2\Phi(\theta_0))$ for $i = 1, \ldots, N$ and $T \geq 3$. In that case $(\sigma_i^2/\widetilde{\sigma}_i^2)$ follows an inverted χ^2 distribution with T degrees of freedom and $E(\sigma_i^2/\widetilde{\sigma}_i^2) = (T-2)^{-1}$ so that $E(\sigma_i/\widetilde{\sigma}_i) < \infty$. It follows from assumption C that when T is fixed, $E(|y_{i,0}\iota'(\Phi(\theta))^{-1}u_i(\theta)/\widetilde{\sigma}_i^2(\theta)|) \leq T\times$ $(\max_{1\leq k\leq T}|(\iota'(\Phi(\theta))^{-1/2})_k|)E(\sigma_i/\widetilde{\sigma}_i(\theta))E(|y_{i,0}/\sigma_i|) < \infty$ and $E(\widetilde{\sigma}_i^{-2}y_{i,0}\iota'(\Phi(\theta_0))^{-1}u_i) = 0$. Furthermore, $E(\widetilde{\sigma}_i^{-2}(\Phi(\theta_0))^{-1/2}u_iu_i'(\Phi(\theta_0))^{-1/2}) = I$. We have the following result:

Theorem 4 *Suppose that $\theta_0 \in \Theta$ and that assumption C holds. Then the RE(Q)MLE and FE(Q)MLE for θ_0 in the model in (29) and the model in (30), respectively, are consistent when $N \to \infty$ irrespective of whether T is fixed or $T \to \infty$.*

The results in theorem 4 still hold when assumption TSH holds and has been imposed on the models in (29) and (30). In this case θ_0 does not include ζ.

When T is fixed or TSH holds and has been imposed on the models, the results follow by applying theorems 2.1 and 2.6 in NMcF; otherwise, the proof of theorem 4 partly follows the proof in Bai (2013), which, instead of SA, uses the stronger assumption SA4, i.e. C(iv). Note that $\text{plim}_{T\to\infty}\widehat{\sigma}_{i,REQMLE}^2 = \sigma_i^2$ and $\text{plim}_{T\to\infty}\widehat{\sigma}_{i,FEQMLE}^2 = \sigma_i^2$.

When TSH holds, assumption C(iii-a) can be replaced by the assumption that the elements of ε_i are symmetrically i.i.d. conditional on v_i and v_i is symmetrically distributed for $i = 1, \ldots, N$.

When $\sigma_i^{-1}u_i$, $i = 1, ..., N$, are i.h.d. and $N, T \to \infty$, a version of theorem 4 holds under

Assumption C$'$:

(i) ε_i, $i = 1, ..., N$, are i.h.d., SA4 with $\varsigma > 0$ and (21) (ii) in case of RE: REA with $\varsigma > 0$, $E(v_i) = 0$, (22) and $v_i \perp y_{i,0}$ (iii) $\exists$ pseudo-true values $\overline{\theta}_0 \in \Theta$ and $\overline{\sigma}_i^2 \in [a, b]$, $i = 1, ..., N$ such that $\overline{\rho} = \rho$, $\overline{\pi} = \widetilde{\pi}$, $\text{plim}_{T\to\infty}\widetilde{\sigma}_i^2(\overline{\theta}_0) = \overline{\sigma}_i^2$, $i = 1, ..., N$,

$\text{plim}_{N,T\to\infty} N^{-1}\sum_{i=1}^{N}\left(\varepsilon_i\varepsilon_i'/\widetilde{\sigma}_i^2(\overline{\theta}_0)\right)=\Psi(\overline{\theta}_0)$, $\text{plim}_{N,T\to\infty} N^{-1}\sum_{i=1}^{N}\left(\varepsilon_i v_i/\widetilde{\sigma}_i^2(\overline{\theta}_0)\right)=0$ and $\text{plim}_{N,T\to\infty} N^{-1}\sum_{i=1}^{N}\left(v_i^2/\widetilde{\sigma}_i^2(\overline{\theta}_0)\right)=\overline{\overline{\sigma}}_v^2$.

Assumption C′(ii) corresponds to a RE formulation of the model but is actually not required for proving large N, T consistency of the REQMLE; assumptions C′(i) and C′(iii) suffice for this purpose, cf. Bai (2013), in particular p.287 and his third lemma on p.291.

Assumption C′ would be useful when the assumption of constant variance ratios, (26) or (27), does not hold, e.g. when the $\sigma_i^{-1}\varepsilon_i$, $i=1,...,N$, are i.i.d. but $\sigma_i^{-1}v_i$, $i=1,...,N$, are i.h.d. Assumption C′ does not require (24), i.e., $E(\varepsilon_{i,t}^2)=\overline{\sigma}_i^2\lambda_t^2$ either. Also, we can have $E(\varepsilon_{i,1}^2)\neq\overline{\sigma}_i^2$. However, assumption C′(iii) implies that $\overline{\zeta}=N^{-1}\sum_{i=1}^{N}\overline{\zeta}_i$ where $diag((E(\varepsilon_{i,1}^2)/\overline{\sigma}_i^2),\overline{\zeta}_i)\equiv Var(\varepsilon_i)/\overline{\sigma}_i^2$, $i=1,...,N$, and $\lim_{T\to\infty}T^{-1}\iota' diag(\overline{\zeta})^{-1}\overline{\zeta}_i=1$, $i=1,...,N$. The latter follows from $\text{plim}_{T\to\infty}\widetilde{\sigma}_i^2(\overline{\theta}_0)=\overline{\sigma}_i^2$, $i=1,...,N$. C′(iii) also implies that $\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}(E(\varepsilon_{i,1}^2)/\overline{\sigma}_i^2)=1$. Thus C′(iii) places non-trivial restrictions on the $E(\varepsilon_{i,t}^2)$ but very mild restrictions on the v_i. The restrictions on $E(\varepsilon_{i,t}^2)$ hold when, for instance, $E(\varepsilon_{i,t}^2)=\sigma_i^2\lambda_t^2\exp(\omega_{i,t})$ where the $\omega_{i,t}$ are i.i.d. with $E(\exp(\omega_{i,t}))=1$. In this case $\overline{\sigma}_i^2=\sigma_i^2$ and $\overline{\zeta}_t=\lambda_t^2$. We have the following result:

Theorem 5 *Suppose that assumption C′ holds. Then the REQMLE and FEQMLE for ρ in the models in (29) and (30), respectively, are consistent when $N\to\infty$ and $T\to\infty$.*

The proof of theorem 5 partly follows Bai (2013). Before we discuss the limiting distributions of the (Q)MLEs, we add the following assumptions:

Assumption D:

If T is fixed, then (a) SA4 and REA4 or FEA4 with $\varsigma=0$ and (b) $E((\sigma_i/\widetilde{\sigma}_i)^2)<\infty$.

We need REA4 or FEA4 to establish the asymptotic distributional properties of the QMLEs when T is fixed, cf. Kruiniger (2013), but not when both $N\to\infty$ and $T\to\infty$, cf. Bai (2013). Letting $G(\theta)=N^{-1}T^{-1}\frac{\partial l(\theta,\widetilde{\sigma}_1^2,...,\widetilde{\sigma}_N^2)}{\partial\theta}\frac{\partial l(\theta,\widetilde{\sigma}_1^2,...,\widetilde{\sigma}_N^2)}{\partial\theta'}$ and $H(\theta)=N^{-1}T^{-1}\frac{\partial^2 l(\theta,\widetilde{\sigma}_1^2,...,\widetilde{\sigma}_N^2)}{\partial\theta\partial\theta'}$, we have:

Theorem 6 *Suppose that $\theta_0\in int(\Theta)$ and that assumptions C and D hold. Then the large N, fixed T asymptotic distribution of the RE(Q)MLE and FE(Q)MLE for θ_0 are given by*

$$\sqrt{N}(\widehat{\theta}_{QMLE}-\theta_0)\xrightarrow{d}N(0,plim_{N\to\infty}(T^{-1}H(\theta_0)^{-1}G(\theta_0)H(\theta_0)^{-1})). \tag{31}$$

The results in theorem 6 follow from theorem 3.4 in NMcF. When $\rho = 1$, the limiting distributions of the QMLEs are non-standard, cf. Kruiniger (2013).

Let $\lambda_{i,t}^2 = \sigma_{i,t}^2/\overline{\sigma}_i^2$, $i = 1, ..., N$, $t = 1, \ldots, T$, $\lambda_1^2 = 1$ and $\lambda_t^2 = N^{-1}\sum_{i=1}^N \lambda_{i,t}^2$, $t = 2, \ldots, T$. Also let $\lambda_1^2 = 1$. Then we have the following result for the REQMLE:

Theorem 7 *Suppose that $\overline{\theta}_0 \in int(\Theta)$ and that assumption C' holds. Let $\omega_T = T^{-1} \times \sum_{t=1}^T \lambda_t^{-2}$, $\overline{\omega} = \lim_{T\to\infty} \omega_T$, $\overline{\gamma} = \lim_{T\to\infty} T^{-1}\sum_{t=2}^T(\lambda_t^{-2}(\lambda_{t-1}^2 + \rho^2\lambda_{t-2}^2 + ... + \rho^{2(t-2)}))$ and $\gamma = \lim_{N,T\to\infty}(NT)^{-1}\sum_{i=1}^N\sum_{t=2}^T(\lambda_t^{-4}\lambda_{i,t}^2(\lambda_{i,t-1}^2 + \rho^2\lambda_{i,t-2}^2 + ... + \rho^{2(t-2)}\lambda_{i,1}^2))$. If $N, T \to \infty$, $T/N^2 \to 0$, $N/T^2 \to 0$ and appropriate Lindeberg conditions hold, then*

$$\sqrt{NT}(\widehat{\rho}_{REQMLE} - \rho) \xrightarrow{d} N\left(0, \gamma/\overline{\gamma}^2\right), \tag{32}$$
$$\sqrt{NT}(\widehat{\widetilde{\pi}}_{REQMLE} - \widetilde{\pi}) \xrightarrow{d} N\left(0, Asyvar(\widehat{\widetilde{\pi}}_{REQMLE})\right),$$
$$\sqrt{NT}(\widehat{\widetilde{\sigma}}_{v,REQMLE}^2 - \widetilde{\sigma}_v^2 - (\frac{1}{N} + \frac{1}{T})\widetilde{b}) \xrightarrow{d} N(0, Asyvar(\widehat{\widetilde{\sigma}}_{v,REQMLE}^2)) \text{ for some } \widetilde{b} \neq 0,$$

and

$$\sqrt{N}(\widehat{\lambda}_{t,REQMLE}^2 - \lambda_t^2) \xrightarrow{d} N(0, Asyvar(\widehat{\lambda}_{t,REQMLE}^2)),\ t = 2, \ldots, T, \text{ and}$$
$$\sqrt{T}(\widehat{\sigma}_{i,REQMLE}^2 - \overline{\sigma}_i^2) \xrightarrow{d} N(0, Asyvar(\widehat{\sigma}_{i,REQMLE}^2)),\ i = 1, \ldots, N. \tag{33}$$

Furthermore, if instead of assumption C', assumption C holds, then $\lambda_{i,t}^2 = \lambda_t^2 = \lambda_t^2$, $t = 1, \ldots, T$ and

$$\sqrt{NT}\begin{bmatrix} \widehat{\rho}_{REQMLE} - \rho \\ \widehat{\widetilde{\sigma}}_{v,REQMLE}^2 - \widetilde{\sigma}_v^2 - (\frac{1}{N} + \frac{1}{T})b \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/\gamma & -\frac{2\widetilde{\sigma}_v^2}{\gamma(1-\rho)} \\ -\frac{2\widetilde{\sigma}_v^2}{\gamma(1-\rho)} & \frac{4\widetilde{\sigma}_v^4}{\gamma(1-\rho)^2} + \frac{4\widetilde{\sigma}_v^2}{\overline{\omega}} \end{bmatrix}\right), \tag{34}$$

where $b = -2(T\omega_T)^{-1}\sum_{t=1}^T(\lambda_t^{-4}\nu_t)(N^{-1}\sum_{i=1}^N(\sigma_i^{-1}v_i))$ with $\nu_t = E(\sigma_i^{-3}\varepsilon_{i,t}^3)$. If in addition $u_i \sim N(0, \sigma_i^2\Phi(\theta_0))$ for $i = 1, \ldots, N$, then $Asyvar(\widehat{\lambda}_{t,REQMLE}^2) = 2\lambda_t^4$, $t = 2, \ldots, T$, and $Asyvar(\widehat{\sigma}_{i,REQMLE}^2) = 2\sigma_i^4$, $i = 1, \ldots, N$, and $\widehat{\rho}$ is a joint N, T asymptotically efficient estimator for ρ. If assumption TSH also holds and TSH has been imposed on the model, $plim_{N,T\to\infty}G(\theta_0) = -plim_{N,T\to\infty}H(\theta_0) = -plim_{N,T\to\infty}N^{-1}T^{-1}\frac{\partial^2 l(\theta_0,\sigma_i^2)}{\partial\theta\partial\theta'}$ and $\sqrt{NT}(\widehat{\theta}_{REQMLE} - \theta_0) \xrightarrow{d} N(0, plim_{N\to\infty}(G(\theta_0))^{-1})$.

A similar theorem can be stated for the FEQMLE. Under random effects with $E(v_i) = 0$, b is actually negligible and $(\frac{1}{N} + \frac{1}{T})b$ can be omitted from the distribution of $\widehat{\widetilde{\sigma}}_{v,REQMLE}^2$.

We only require $N/T^3 \to 0$ for the joint N,T limiting distribution of $\widehat{\rho}_{REQMLE}$ to have the simple form in Theorem 7, cf. Bai (2013). The condition $T/N^2 \to 0$ is used only for the joint N,T limiting distribution of $\widehat{\widetilde{\sigma}}^2_{v,REQMLE}$, while $N/T^2 \to 0$ is used for the limiting distributions of $\widehat{\lambda}^2_{t,REQMLE}$, $t = 2,\ldots,T$, cf. Bai (2013).

When assumptions C and TSH hold and TSH has been imposed on the model, the results in theorem 7 follow from theorem 3.2 or 3.4 in NMcF; otherwise we can follow Bai (2013). Note that if assumption TSH holds, $\gamma = \overline{\gamma} = (1-\rho^2)^{-1}$.

One can obtain a consistent estimator for the asymptotic variance of the weighted QMLE for θ when $E(\varepsilon^2_{i,t}) = \sigma^2_i\lambda^2_t$ (or for ρ if $E(\varepsilon^2_{i,t}) \neq \sigma^2_i\lambda^2_t$) by using a subsampling method. Hayakawa and Pesaran (2012) derive a consistent estimator for the large N, fixed T asymptotic variance of the unweighted FEMLE for θ when $E(u_iu_i') = \sigma^2_i(\widetilde{\sigma}^2_v\iota\iota' + I)$.

When assumptions C or C$'$(iii) do not hold, one may still be able to obtain a joint N,T consistent estimator for ρ that is asymptotically more efficient than the unweighted QMLE for ρ by following a two-step procedure: let $\overline{\zeta} = (\lambda^2_1, \lambda^2_2 \ldots \lambda^2_T)'$, redefine $\theta_0 = (\rho, \tilde{\pi}, \widetilde{\sigma}^2_v, \overline{\zeta}')'$ and $\Phi(\theta_0) = \widetilde{\sigma}^2_v\iota\iota' + diag(\overline{\zeta})$ accordingly, and use the unweighted Quasi MLE for θ, say $\widehat{\theta}_U$, to compute the estimates $\widetilde{\sigma}^2_i(\widehat{\theta}_U) = T^{-1}tr([\Phi(\widehat{\theta}_U)]^{-1}u_i(\widehat{\theta}_U)u_i'(\widehat{\theta}_U))$ for the σ^2_i. Next replace s^2_i in the redefined weighted log-likelihood function $l(\theta, s^2_1, ..., s^2_N)$ by $\widetilde{\sigma}^2_i(\widehat{\theta}_U)$ $\forall i \in \{1,...,N\}$ and maximize the resulting function w.r.t. θ to obtain the weighted Quasi MLE for ρ (and π). Under appropriate (moment) conditions joint N,T consistency of this weighted QMLE for ρ can be shown similarly as joint N,T consistency of the unweighted QMLE for ρ by partly following Bai (2013).

Let $\widetilde{\sigma}^2_{i,t} = w^2_i\sigma^2_{i,t}$ and $\bar{\sigma}^2_{w,t} = N^{-1}\sum_{i=1}^N \widetilde{\sigma}^2_{i,t}$, $i = 1,...,N$, $t = 1,\ldots,T$, where the weights w^2_i may be chosen as e.g. $w^2_i = \widetilde{\sigma}_i^{-2}(\widehat{\theta}_U)$, $i = 1,...,N$. Then we have the following result on the joint N,T distribution of the weighted REQMLE for ρ, $\widehat{\rho}_{REWQMLE}$:

Theorem 8 *Suppose that $\theta_0 \in int(\Theta)$ and that assumptions $C'(i)$ and $C'(ii)$ hold. Let $w^2_i = \widetilde{\sigma}_i^{-2}(\widehat{\theta}_U)$, $i = 1,...,N$, $\widetilde{\overline{\gamma}} = \lim_{T\to\infty} T^{-1}\sum_{t=2}^T(\bar{\sigma}^{-2}_{w,t}(\bar{\sigma}^2_{w,t-1}+\rho^2\bar{\sigma}^2_{w,t-2}+...+\rho^{2(t-2)}\bar{\sigma}^2_{w,1}))$ and $\widetilde{\gamma} = \lim_{N,T\to\infty}(NT)^{-1}\sum_{i=1}^N\sum_{t=2}^T(\bar{\sigma}^{-4}_{w,t}\widetilde{\sigma}^2_{i,t}(\widetilde{\sigma}^2_{i,t-1}+\rho^2\widetilde{\sigma}^2_{i,t-2}+...+\rho^{2(t-2)}\widetilde{\sigma}^2_{i,1}))$. If $N,T \to \infty$, $N/T^3 \to 0$, $\widehat{\rho}_{REWQMLE}$ is consistent and appropriate Lindeberg conditions are satisfied, then $\sqrt{NT}(\widehat{\rho}_{REWQMLE} - \rho) \xrightarrow{d} N\left(0, \widetilde{\gamma}/(\widetilde{\overline{\gamma}})^2\right)$.*

A similar result holds for other choices of w^2_i for which $\widehat{\rho}_{REWQMLE}$ is consistent and appropriate Lindeberg conditions are satisfied. The unweighted REQMLE for ρ, $\widehat{\rho}_{REUQMLE}$, is obtained for $w^2_i = 1$, $i = 1,...,N$.

When $E(\varepsilon_{i,t}^2) = \sigma_i^2\lambda_t^2$, $N, T \to \infty$, $N/T^3 \to 0$ and appropriate Lindeberg conditions hold, then Theorem 8 implies that $Asyvar(\widehat{\rho}_{REWQMLE}) = (\widetilde{\overline{\gamma}})^{-1}(N^{-1}\sum_{i=1}^N w_i^4\sigma_i^4)\times (N^{-1}\sum_{i=1}^N w_i^2\sigma_i^2)^{-2}$ and hence $Asyvar(\widehat{\rho}_{REUQMLE})/Asyvar(\widehat{\rho}_{REQMLE}) = (N^{-1}\sum_{i=1}^N \sigma_i^4)\times (N^{-1}\sum_{i=1}^N \sigma_i^2)^{-2} > 1$ if $\sigma_i^2 \neq \sigma_j^2$ for some $i, j \in \{1, ..., N\}$. In practice, if it turns out that the weighted QMLE for ρ is not more efficient than the unweighted QMLE for ρ, then one can simply ignore the weighted QMLE.

By repeating the second step of the two-step weighted QMLE described above one can obtain an iterated weighted QMLE. At each iteration the weights $\widetilde{\sigma}_i^2(\widehat{\theta}) = T^{-1}tr([\Phi(\widehat{\theta})]^{-1}\times u_i(\widehat{\theta})u_i'(\widehat{\theta}))$, $1, ..., N$, are updated by substituting the latest estimate of θ for $\widehat{\theta}$ in $\widetilde{\sigma}_i^2(\widehat{\theta})$, $1, ..., N$. Instead of $\widetilde{\sigma}_i^2(\widehat{\theta})$, one may use normalised weights $\overline{\zeta}_1\widetilde{\sigma}_i^2(\widehat{\theta})$, $1, ..., N$. This is equivalent to dividing the variance parameters by $\overline{\zeta}_1$. If the $\sigma_{i,t}^2 = E(\varepsilon_{i,t}^2)$ satisfy (24) or C$'$(iii), then the iterated estimates of $\overline{\zeta}_1$ will converge to 1 when using the normalised weights.

Finally, we describe a hybrid QMLE that is based on an objective function that nests the weighted likelihood function but is also a quasi likelihood function. Recall that $z = (l_2^2 \ ... \ l_T^2)'$, $\Phi(\widetilde{s}_v^2, z) = \widetilde{s}_v^2\iota\iota' + \Psi(z)$ and $u_i(\theta) = y_i - ry_{i,-1} - \widetilde{p}y_{i,0}\iota$. Let $\bar{z} = (\bar{l}_1^2, \ \bar{l}_2^2 \ ... \ \bar{l}_T^2)'$, $\overline{\Phi}(\bar{s}_v^2, \bar{z}) = \bar{s}_v^2\iota\iota' + diag(\bar{z})$ and redefine $\theta = (r, \widetilde{p}, \bar{s}_v^2, \widetilde{s}_v^2, \bar{z}', z')'$. Then the 'hybrid' objective function is $\bar{l}(\theta) = \sum_{i=1}^N \bar{l}_{RE,i}(\theta)$ with

$$\begin{aligned}\bar{l}_{RE,i}(\theta) &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\overline{\Phi}(\bar{s}_v^2, \bar{z})\right| - \frac{T}{2}\ln\widetilde{\sigma}_i^2(\theta) - \\ &\quad \frac{1}{2\widetilde{\sigma}_i^2(\theta)}(u_i(\theta))'(\overline{\Phi}(\bar{s}_v^2, \bar{z}))^{-1}u_i(\theta),\end{aligned}$$

where $\widetilde{\sigma}_i^2(\theta) = T^{-1}tr([\Phi(\widetilde{s}_v^2, z)]^{-1}u_i(\theta)u_i'(\theta))$.

We make the following remarks. The hybrid QMLE for ρ will be joint N, T asymptotically equivalent to the weighted QMLE for ρ when the $\sigma_{i,t}^2 = E(\varepsilon_{i,t}^2)$ satisfy (24) or C$'$(iii). In that case, one would expect that the estimates for $\overline{\zeta} - (1, \zeta)'$ are not significantly different from zero. The hybrid QMLE for ρ will be large N, fixed T asymptotically equivalent to the weighted QMLE for ρ when assumption C holds, in particular, when (24) and either (26) or (27) are satisfied. However, unlike the weighted QMLE, the hybrid QMLE will generally still be consistent when the $\sigma_{i,t}^2 = E(\varepsilon_{i,t}^2)$ do not satisfy (24) or C$'$(iii). In that case the hybrid QMLE, like the iterated QMLE, may still be more efficient than the unweighted QMLE. The hybrid QMLE does not require iteration and is also more flexible than the iterated QMLE. For instance, it is possible that $\overline{\zeta} \neq a(1, \zeta)'$ for all $a \in \mathbb{R}^+$.

4.3 A Panel GARCH(1,1) model

In this section we briefly discuss ML estimation of panel GARCH models. Such models have been considered before by Engle and Mezrich (1996), Bauwens and Rombouts (2007) and Pakel, Shephard and Sheppard (2011). Similarly to Pakel et al. (2011) we model the conditional variance of the idiosyncratic error, i.e., $E(\varepsilon_{i,t}^2|\mathfrak{I}_{t-1}) \equiv h_{i,t}$, where $\mathfrak{I}_{t-1}$ is a set containing information up to period $t-1$, as

$$h_{i,t} = \sigma_i^2(1-\alpha_1-\beta_1) + \alpha_1\varepsilon_{i,t-1}^2 + \beta_1 h_{i,t-1}$$

with $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\alpha_1+\beta_1 < 1$. This model for $E(\varepsilon_{i,t}^2|\mathfrak{I}_{t-1})$ implies that $E(\varepsilon_{i,t}^2) = \sigma_i^2\tau_t^2$ with $\tau_t^2 = 1,\ t = 1,...,T$.

When N is large but T is fixed, one cannot consistently estimate the GARCH parameters α_1, β_1 and σ_i^2, $i = 1,...,N$, due to the combination of the presence of the incidental variance parameters and the fact that the $h_{i,t}$ (and the $\varepsilon_{i,t}^2$) $i = 1,...,N$, $t = 0,...,T-1$ are unobserved. However, when T grows large and $\alpha_1+\beta_1 < 1$, the effect of the choices for the values of $h_{i,0}$ $i = 1,...,N$ on the estimator becomes negligible. Furthermore, in case the mean equations of the panel model contain incidental parameters, η_i, $i = 1,...,N$ (cf. the panel AR(1) model), one can obtain consistent estimates of such parameters when T grows large. This is important because $h_{i,t}$ depends on the η_i through the $\varepsilon_{i,t-1}^2$ terms.

Thus when $N, T \to \infty$, one can consistently estimate all the model parameters. In this case one can take the following approach to ML estimation of the model. First estimate the σ_i^2 (and the η_i if present) using a likelihood function that ignores the GARCH specification but uses the unconditional variances, i.e., the σ_i^2, instead. Then in the next step maximize the likelihood function that includes the GARCH specification using the $\widehat{\sigma}_i^2$ and the $\widehat{\eta}_i$ as starting values for the σ_i^2 and the η_i. Note that the m-profile composite likelihood estimator of Pakel et al. (2011) also uses a similar first step but in the second step keeps the values of the σ_i^2 fixed at their initial estimates. Hence their estimator will generally be less efficient than the standard ML estimator.

5 GMM estimation under multiplicative cross-sectional heteroskedasticity

5.1 Static panel data models

Suppose that $Dy_i \sim N(0, \sigma_i^2 D\Phi(\theta)D')$. Note that this model is a simplified version of the model from section 2 with $\Phi_i(\theta) = \Phi(\theta)$, $i = 1, ..., N$, and in particular that $\Phi(\theta)$ does not depend on W_i. Let $\widehat{\theta}_{FD} = \widehat{\theta}_{FDMLE}$ and $\widetilde{\sigma}_i^2(\theta) = \frac{1}{T-1}tr([\Phi(\theta)]^{-1}Z_i)$, $i = 1, ..., N$, (cf. (6)). The FDMLE for θ is equivalent to a GMM estimator that exploits the moment conditions in

$$E[([\Phi(\widehat{\theta}_{FD})]^{-1/2} \otimes [\Phi(\widehat{\theta}_{FD})]^{-1/2})vec(\Phi(\theta) - Z_i/\widetilde{\sigma}_i^2(\widehat{\theta}_{FD}))] = 0 \tag{35}$$

and uses the identity matrix as weight matrix. Clearly, $\widehat{\theta}_{FD}$ is also equivalent to a GMM estimator that exploits the moment conditions in

$$E[vech([\Phi(\widehat{\theta}_{FD})]^{-1/2}[\Phi(\theta) - Z_i/\widetilde{\sigma}_i^2(\widehat{\theta}_{FD})][\Phi(\widehat{\theta}_{FD})]^{-1/2})] = 0. \tag{36}$$

Note that $S_i(\theta) = [\Phi(\theta)]^{-1/2}Z_i[\Phi(\theta)]^{-1/2}/((T-1)\widetilde{\sigma}_i^2(\theta))$. An optimal weight matrix for (36) is $[T^2Cov(vech(S_i(\theta_0)))]^{-1}$, which is a diagonal matrix. Using well known results about the moments of the Dirichlet distribution, which is a multivariate generalisation of the Beta distribution, see Johnson and Kotz (1972), it can be shown that the diagonal elements of $Cov(vech(S_i(\theta_0)))$ that correspond to the diagonal elements of $S_i(\theta_0)$ are equal to $2(T-2)/((T-1)(T^2-1))$ while its other diagonal elements are equal to $1/(T^2-1)$. It follows that the optimal weight matrix for (35) is only proportional to the identity matrix when $T \to \infty$ but not when T is fixed. We conclude that when T is fixed the GMM estimator that exploits (36) and uses $[T^2Cov(vech(S_i(\theta_0)))]^{-1}$ as weight matrix is asymptotically more efficient than $\widehat{\theta}_{FD}$. The fixed T, large N asymptotic distribution of this *Individually Weighted* (IW) GMM estimator is given by:

$$\sqrt{N}(\widehat{\theta}_{IWGMM} - \theta) \xrightarrow{d} N(0, (E(\frac{dvechS_i(\theta)}{d\theta'}|_{\theta_0})'Cov(vechS_i(\theta_0))^{-1}E(\frac{dvechS_i(\theta)}{d\theta'}|_{\theta_0}))^{-1}). \tag{37}$$

The efficiency loss of the FDMLE is related to the presence of incidental variance parameters.

Note that the FDMLE for θ is also large N asymptotically equivalent to a GMM estimator for θ that exploits

$$
\begin{aligned}
&E[vech(\Phi(\theta) - N^{-1}\textstyle\sum_{i=1}^{N}(Z_i/s_i^2))] = 0, \\
&E[s_i^2 - (T-1)^{-1}tr([\Phi(\theta)]^{-1}Z_i)], \quad i = 1, ..., N.
\end{aligned} \tag{38}
$$

A notable feature of the moment conditions in (35), (36) and (38) is that the individual observations $Z_i = Dy_iy_i'D'$ are weighted by estimates of $1/\sigma_i^2$, $i = 1, ..., N$. This weighting leads to an estimator that uses i.i.d. data. Furthermore, when there is a considerable degree of heteroskedasticity in the cross-section dimension, this type of weighting potentially increases the precision of the resulting Optimal GMM estimator substantially as compared to the precision of the Optimal GMM estimators that are based on

$$E(\widetilde{\sigma}_i^2(\theta)\Phi(\theta) - Z_i) = 0 \tag{39}$$

or $E(\sigma^2\Phi(\theta) - Z_i) = 0$, respectively. On the other hand, the fact that the Z_i are weighted by estimates of $1/\sigma_i^2$ rather than by the true values of $1/\sigma_i^2$, contributes to the variance of the (Optimal) GMM estimator that exploits (35), (36) or (38). Therefore, when T is small, the Optimal GMM estimator that exploits (35), (36) or (38) could actually be less efficient than the Optimal GMM estimator that exploits (39).

The Individually Weighted GMM estimator that exploits (35), (36) or (38) can be thought of as a generalization of the Minimum Distance estimators of Chamberlain (1982, 1984).

5.2 Nonstationary panel AR(1) model

We consider GMM estimation of the nonstationary panel AR(1) model in (16) under assumption C.

Let $w_{i,t}(r) = y_{i,t} - ry_{i,t-1} = u_{i,t}(\theta) + \widetilde{p}y_{i,0}$ and let Δ be the first-difference operator so that $\Delta w_{i,t}(r) = w_{i,t}(r) - w_{i,t-1}(r) = \Delta u_{i,t}(\theta)$. Ahn & Schmidt (1995) have shown that under assumptions C and REA and CS homoskedasticity a large N, fixed T asymptotically efficient GMM estimator for ρ in (16) exploits the following moment conditions:

$$\begin{aligned} E(y_{i,0}\Delta w_{i,t}) &= E(y_{i,0}\Delta u_{i,t}) = 0, \quad t = 2, ..., T, \\ E(w_{i,s}\Delta w_{i,t}) &= E(u_{i,s}\Delta u_{i,t}) = 0, \quad t = 3, ..., T, s = 1, ..., t-2, \\ E(w_{i,T}\Delta w_{i,t}) &= E(u_{i,T}\Delta u_{i,t}) = 0, \quad t = 2, ..., T-1. \end{aligned} \tag{40}$$

Under the same assumptions and normality the Optimal GMM estimator exploiting (40) is large N, fixed T asymptotically equivalent to the unweighted REML estimator for ρ in (16), cf. Kruiniger (2013). When the errors are CS heteroskedastic, the Weighted REMLE for ρ proposed in section 4 may be more efficient than these unweighted estimators. Instead of this Weighted REML estimator, one may use an Individually Weighted GMM estimator that exploits the following moment conditions:

$$\begin{aligned} E(\widetilde{\sigma}_i^{-2}(\widehat{\theta})y_{i,0}\Delta w_{i,t}) &= 0, \quad t = 2, ..., T, \\ E(\widetilde{\sigma}_i^{-2}(\widehat{\theta})w_{i,s}\Delta w_{i,t}) &= 0, \quad t = 3, ..., T, \ s = 1, ..., t-2, \\ E(\widetilde{\sigma}_i^{-2}(\widehat{\theta})w_{i,T}\Delta w_{i,t}) &= 0, \quad t = 2, ..., T-1, \end{aligned} \tag{41}$$

where $\widetilde{\sigma}_i^{-2}(\theta)$ is defined above assumption C and $\widehat{\theta}$ is an initial consistent estimator of θ.

The Weighted REMLE for ρ is only large N, fixed T asymptotically efficient under normality of the data, whereas the Optimal IW GMM estimator based on (41) will also be large N, fixed T asymptotically efficient under non-normality. The Weighted REMLEs for $\widetilde{\sigma}_v^2$, λ_2^2, ..., λ_T^2 are large N, fixed T asymptotically inefficient even under normality. The reasons for the inefficiency of these estimators under normality are similar to those given in section 5.1 for the large N, fixed T asymptotic inefficiency of the Weighted FDMLEs for the parameters in θ and are related to the presence of incidental variance parameters.

Note that like the Weighted REMLE, consistency of the GMM estimator based on (41) requires symmetry of the distributions of the elements of $(\Phi(\theta_0))^{-1/2}u_i$. An alternative Individually Weighted GMM estimator for ρ that does not require this assumption is based on the following moment conditions:

$$\begin{aligned} E(\widetilde{\sigma}_{i,-(t-1),-t}^{-2}(\widehat{\theta})y_{i,0}\Delta w_{i,t}) &= 0, \quad t = 2, ..., T, \\ E(\widetilde{\sigma}_{i,-s,-(t-1),-t}^{-2}(\widehat{\theta})w_{i,s}\Delta w_{i,t}) &= 0, \quad t = 3, ..., T, \ s = 1, ..., t-2, \\ E(\widetilde{\sigma}_{i,-(t-1),-t,-T}^{-2}(\widehat{\theta})w_{i,T}\Delta w_{i,t}) &= 0, \quad t = 2, ..., T-1, \end{aligned} \tag{42}$$

where

$$\widetilde{\sigma}^2_{i,-(t-1),-t}(\theta) = \frac{1}{T-2} tr([\Phi_{-(t-1),-t}(\theta)]^{-1} u_{i,-(t-1),-t}(\theta) u'_{i,-(t-1),-t}(\theta)),$$
$$\Phi_{-(t-1),-t}(\theta) = \widetilde{\sigma}^2_v \iota_{T-3} \iota'_{T-3} + diag(1, \lambda^2_2, \lambda^2_3, ..., \lambda^2_{t-2}, \lambda^2_{t+1}, ..., \lambda^2_T),$$
$$u_{i,-(t-1),-t}(\theta) = (u_{i,1}(\theta), u_{i,2}(\theta), ..., u_{i,t-2}(\theta), u_{i,t+1}(\theta), ..., u_{i,T}(\theta)),$$

$\widetilde{\sigma}^{-2}_{i,-(t-1),-t,-T}(\theta)$ is defined similarly, and

$$\widetilde{\sigma}^2_{i,-s,-(t-1),-t}(\theta) = \frac{1}{T-3} tr([\Phi_{-s,-(t-1),-t}(\theta)]^{-1} u_{i,-s,-(t-1),-t}(\theta) u'_{i,-s,-(t-1),-t}(\theta)),$$
$$\Phi_{-s,-(t-1),-t}(\theta) = \widetilde{\sigma}^2_v \iota_{T-3} \iota'_{T-3} + diag(1, \lambda^2_2, \lambda^2_3, ..., \lambda^2_{s-1}, \lambda^2_{s+1}, ..., \lambda^2_{t-2}, \lambda^2_{t+1}, ..., \lambda^2_T),$$
$$u_{i,-s,-(t-1),-t}(\theta) = (u_{i,1}(\theta), u_{i,2}(\theta), ..., u_{i,s-1}(\theta), u_{i,s+1}(\theta), ..., u_{i,t-2}(\theta), u_{i,t+1}(\theta), ..., u_{i,T}(\theta)).$$

To reduce the computational burden, $E(\widetilde{\sigma}^{-2}_{i,-(t-1),-t}(\widehat{\theta}) y_{i,0} \Delta w_{i,t}) = 0, \quad t = 2, ..., T$, can be replaced by $E(\widetilde{\sigma}^{-2}_{i,-s,-(t-1),-t}(\widehat{\theta}) y_{i,0} \Delta w_{i,t}) = 0, \quad t = 2, ..., T$, for some $s \in \{1, ..., t-2\}$.

6 Monte Carlo results

In this section we compare through Monte Carlo simulations the finite sample properties of the unweighted RE- and FEQMLE and the weighted RE- and FEQMLE for ρ. We study how the properties of these estimators are affected if we change (1) the skedastic properties of the idiosyncratic errors (the $\varepsilon_{i,t}$), (2) the distributions of the $v_{i,0} = y_{i,0} - \mu_i$ or (3) in the case of the REQMLEs the ratios of the variances of the error components, i.e. $\sigma^2_{\mu,i}/\sigma^2_i$. We conducted the simulation experiments for $(T, N) = (4, 100)$, $(4, 500)$, $(9, 100)$, $(9, 500)$ or $(24, 100)$ and $\rho = 0.2$, 0.5, 0.8 or 0.95.

In all simulation experiments the error components have been drawn from normal distributions with zero means. We assumed that $\sigma^2_{\mu,i}/\sigma^2_i = 0$, 1 or 25. For the variances of the $\varepsilon_{i,t}$ we considered three different designs (identified by a Roman number):

I Homoskedasticity of the $\varepsilon_{i,t}$: $E(\varepsilon^2_{i,t}) = \sigma^2$ with $\sigma^2 = 1$.

II CS heteroskedasticity of $\varepsilon_{i,t}$: $E(\varepsilon^2_{i,t}) = \sigma^2_i$ with $\sigma^2_i \sim Uniform[0.4, 1.6]$.

III CS heteroskedasticity of $\varepsilon_{i,t}$: $E(\varepsilon^2_{i,t}) = \sigma^2_i$ with $\sigma^2_i \sim \chi^2(1)$.

IV CS heteroskedasticity of $\varepsilon_{i,t}$: $\varepsilon_{i,t} \sim \sigma_i(\chi^2(1) - 1)/\sqrt{2}$ with $\sigma^2_i \sim \chi^2(1)$.

V Time-series heteroskedasticity of $\varepsilon_{i,t}$: $E(\varepsilon_{i,t}^2) = \sigma_{i,t}^2$ with $\sigma_{i,t}^2 = 0.4 + 0.8(t-1)/(T-1) + 0.4U_{i,t}$, where $U_{i,t} \sim uniform[0,1]$, $t = 1, ..., T$.

VI Interactive heteroskedasticity of $\varepsilon_{i,t}$: $E(\varepsilon_{i,t}^2) = \sigma_{i,t}^2$ with $\sigma_{i,t}^2 = \sigma_i^2(0.4+0.8(t-1)/(T-1) + 0.4U_{i,t})$, where $\sigma_i^2 \sim \chi^2(1)$ and $U_{i,t} \sim uniform[0,1]$, $t = 1, ..., T$.

We also assumed cross-sectional independence and absence of autocorrelation in the $\varepsilon_{i,t}$.

In order to assess how the distributions of the $v_{i,0} = y_{i,0} - \mu_i$ affect the properties of the estimators, we conducted two different sets of experiments (identified by a capital): in one set, labeled NS, the initial observations are non-stationary, i.e., $y_{i,0} - \mu_i = 0$, $i = 1, ..., N$, whereas in the other set, labeled S, the initial observations are drawn from stationary distributions, i.e., $(y_{i,0} - \mu_i) \sim N(0, \sigma_i^2/(1-\rho^2))$, $i = 1, ..., N$.

Note that all four estimators suffer from a weak moment conditions problem under time series homoskedasticity when ρ is close to one. Furthermore, $\widehat{\rho}_{FEQML}$ is a constrained version of $\widehat{\rho}_{REQML}$, cf. Kruiniger (2013).

When we conducted the experiments, we did not impose time series homoskedasticity on the likelihood functions. However, we did add the restrictions $l_t^2 > 0$ and $(T-1)\widetilde{s}_v^2 + l_t^2 > 0$, $t = 1, ..., T$, to the likelihood functions – with $l_1^2 = 1$ in the case of the weighted QMLE – to ensure that the estimates of $E(u_i u_i')$ were positive definite. We also allowed for time effects by subtracting cross-sectional averages from the data.

Tables 1-18 report the simulation results in terms of the biases and root mean squared errors (RMSEs) of the four QML estimators for ρ. The tables differ with respect to the assumptions made about the skedastic properties of the $\varepsilon_{i,t}$, the distributions of the $v_{i,0} = y_{i,0} - \mu_i$ and the value of N. In all these tables $\sigma_{\mu,i}^2/\sigma_i^2 = 1$. Tables 1 and 2 correspond to design I-S, tables 3 and 4 corresponds to design II-S, tables 5 and 6 correspond to design III-S, tables 7 and 8 correspond to design I-NS, tables 9 and 10 correspond to design III-NS, tables 11 and 12 correspond to design IV-S, tables 13 and 14 corresponds to design IV-NS, tables 15 and 16 correspond to design V-S, and tables 17 and 18 correspond to design VI-S In the tables the unweighted QMLEs are labeled RE- and FEUQMLE whereas the weighted estimators are labeled as RE- and FEWQMLE.

Inspection of the results in tables 1-18 leads to the following conclusions:

1. Under designs I-S, I-NS and V-S (i.e., CS homoskedasticity) the unweighted QMLEs are more efficient than their weighted counterparts. The inefficiency of the weighted

QMLEs decreases when T increases.

2. Under design II-S the weighted QMLEs are more efficient than their unweighted counterparts when $T = 24$ and $N = 100$.

3. Under designs III-S.and III-NS the weighted QMLEs are more efficient than their unweighted counterparts when $T = 9$ or $T = 24$ and the FEWQMLE is also more efficient than the FEUQMLE when $T = 4$ and $\rho > 0.5$.

4. Under designs IV-S.and IV-NS the weighted QMLEs are only more efficient than their unweighted counterparts when $T = 9$ or $T = 24$ and $\rho = 0.95$.

5. Under designs VI-S.and VI-NS the REWQMLEs are more efficient than the REUQMLEs when $T = 9$ or $T = 24$ and the FEWQMLEs are also more efficient than the FEUQMLEs when $T = 9$ and $\rho = 0.2$ or 0.5 or $T = 24$.

6. Under designs S the RE estimators are generally more efficient than their FE counterparts. Moreover, under design III-S the REWQMLE is the most efficient estimator when $T = 4$ and $\rho = 0.8$.

7. Under designs NS the RMSEs of the RE estimators are similar to the RMSEs of their FE counterparts when $T = 9$ or $T = 24$ but under designs I-NS and III-NS the FE estimators are more efficient than their RE counterparts when $T = 4$.

7 Conclusions

In this paper we have proposed asymptotically efficient ML estimators and GMM estimators for the common parameters in panel data models with (multiplicative) fixed effects in the variance, e.g. $\sigma_{i,t}^2 = \sigma_i^2$ or more generally $\sigma_{i,t}^2 = \sigma_i^2 \lambda_t^2$. We considered estimation of both 'static' models, including covariance stationary models, and dynamic models with arbitrary initial conditions, i.e., 'non-stationary' models.

A notable feature of the likelihood equations and the moment conditions for models with (multiplicative) fixed variance effects is the weighting of the data by the individual variances. This weighting may lead to a considerable improvement in precision of the corresponding ML and GMM estimators.

The theory in sections 2 and 4 suggests that the FEML estimator for β in the model $y_i = \mu_i \iota + X_i \beta + \varepsilon_i$, $i = 1, ..., N$, where $\{(\sigma_i^{-1} Q X_i, \sigma_i^{-1} \varepsilon_i)\}$ is an $i.i.d.$ sequence with $\sigma_i^{-1} \varepsilon_i \sim$

$N(0, I)$, $E(X_i' Q \varepsilon_i) = 0$ and $Q = I - T^{-1}\iota\iota'$, is consistent under suitable conditions and either large N asymptotics or large N, T asymptotics. Normality of ε_i is not necessary for consistency of this estimator: when T is fixed we require symmetry of the distributions of the ε_i and $E(\sup_b(\sigma_i/\widetilde{\sigma}_i(b))) < \infty$ where $\widetilde{\sigma}_i^2(b) = \frac{1}{T-1}(u_i'(b) Q u_i(b))$ with $u_i(b) = y_i - X_i b$, $i = 1, ..., N$, cf. assumption C(iii) in section 4. We also require that $E(T^{-1}\sigma_i^{-2} X_i' Q X_i)$ is finite (for all T). Identification of β requires that $E(T^{-1}\sigma_i^{-2} X_i' Q X_i)$ is nonsingular (for all T). The FEML estimator for β can be reinterpreted as the (iterated) Weighted Least Squares Dummy Variables (LSDV) estimator.

The stationary dynamic panel data model discussed in MaCurdy (1982) and Ahn and Schmidt (1997) fits the framework discussed in the paper. However, the idea of individual (variance) weighted moment conditions can also be applied to Arellano-Bond (1991) type moment conditions which are valid under completely arbitrary heteroskedasticity in both dimensions and arbitrary initial conditions. For instance, if T is not too small and there is a substantial degree of heteroskedasticity in the cross-section dimension, then a moment condition like $E((\Delta y_{i,t} - \rho \Delta y_{i,t-1}) y_{i,s} / (\sum_{k=2}^{T} (\Delta y_{i,k} - \rho \Delta y_{i,k-1})^2)) = 0$ with $s \leq t - 2$ may lead to a more efficient GMM estimator for the autoregressive parameter ρ than $E((\Delta y_{i,t} - \rho \Delta y_{i,t-1}) y_{i,s}) = 0$ with $s \leq t - 2$.

A Proof of a result in section 2

Recall that $H(\theta) = E(T^{-1}\frac{dg_i(\theta)}{d\theta'})$, where $g_i(\theta) = -(\frac{dvec\Phi_i}{d\theta'})'([\Phi_i(\theta)]^{-1} \otimes [\Phi_i(\theta)]^{-1})\times vec(\Phi_i(\theta) - Z_i/\widetilde{\sigma}_i^2(\theta)) = -F_i(\theta)vec[I - (T-1)S_i(\theta)]$.

We will show that $H(\theta_0) = -T^{-1}E\{F_i(\theta_0)[I - (T-1)(vecS_i(\theta_0)(vecS_i(\theta_0))')]F_i'(\theta_0)\}$. In the derivation we will use two important results from linear algebra: for any conformable matrices A, B, and C, we have $vec(ABC) = (C' \otimes A)vecB$ and $trAB = (vecA')'vecB$ (see Magnus and Neudecker (1988)).

Noting that $(T-1)\widetilde{\sigma}_i^2(\theta) = tr([\Phi_i(\theta)]^{-1}Z_i)$, we have

$$(T-1)\frac{d\widetilde{\sigma}_i^2(\theta)}{d\theta'} = \frac{(vecZ_i)'(dvec[\Phi_i(\theta)]^{-1})}{d\theta'}. \tag{43}$$

After rewriting $(vecZ_i)'$ as $(vec(\Phi_i^{1/2}\Phi_i^{-1/2}Z_i\Phi_i^{-1/2}\Phi_i^{1/2}))' = (vec(\Phi_i^{-1/2}Z_i\Phi_i^{-1/2}))' \times (\Phi_i^{1/2} \otimes \Phi_i^{1/2})$, and using that

$$\frac{dvec[\Phi_i(\theta)]^{-1}}{d\theta'} = -(\Phi_i^{-1} \otimes \Phi_i^{-1})\frac{dvec[\Phi_i(\theta)]}{d\theta'}, \tag{44}$$

which is a standard result on matrix differentiation (see Magnus and Neudecker (1988)), we obtain

$$\begin{aligned}(T-1)\frac{d\widetilde{\sigma}_i^2(\theta)}{d\theta'} &= -(vec(\Phi_i^{-1/2}Z_i\Phi_i^{-1/2}))'(\Phi_i^{-1/2} \otimes \Phi_i^{-1/2})\frac{dvec\Phi_i(\theta)}{d\theta'} = \\ &\quad -(T-1)\widetilde{\sigma}_i^2(\theta)[vecS_i(\theta)]'F_i'(\theta). \end{aligned} \tag{45}$$

Recalling that $E(\Phi_i(\theta_0) - Z_i/\widetilde{\sigma}_i^2(\theta_0)) = 0$, the result now follows straightforwardly.

B Proof of results in section 3

We give the proofs only for the RE(Q)MLE. The proofs for the FE(Q)MLE are similar. We first derive the score vector $\frac{\partial l}{\partial \theta_0}$, where $\theta_0 = (\rho, \tilde{\pi}, \widetilde{\sigma}_v^2, \zeta')'$ and $l = \sum_{i=1}^N l_{RE,i}(\theta_0, \sigma_i^2)$ with

$$\begin{aligned} l_{RE,i}(\theta_0, \sigma_i^2) &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\Phi| - \frac{T}{2}\ln\sigma_i^2 - \\ &\quad \frac{1}{2\sigma_i^2}(y_i - \rho y_{i,-1} - \tilde{\pi}y_{i,0}\iota)'\Phi^{-1}(y_i - \rho y_{i,-1} - \tilde{\pi}y_{i,0}\iota), \end{aligned}$$

where $\Phi = \Phi(\theta_0)$. We obtain

$$\begin{aligned}
\frac{\partial l}{\partial \rho} &= \sum_{i=1}^{N} (\frac{1}{\sigma_i^2} y_{i,-1}' \Phi^{-1}(y_i - \rho y_{i,-1} - \tilde{\pi} y_{i,0}\iota)), \\
\frac{\partial l}{\partial \tilde{\pi}} &= \sum_{i=1}^{N} (\frac{y_{i,0}}{\sigma_i^2} \iota' \Phi^{-1}(y_i - \rho y_{i,-1} - \tilde{\pi} y_{i,0}\iota)), \\
\frac{\partial l}{\partial \widetilde{\sigma}_v^2} &= \frac{1}{2} \sum_{i=1}^{N} tr(\frac{d\Phi^{-1}}{\partial \widetilde{\sigma}_v^2}(\Phi - \frac{u_i u_i'}{\sigma_i^2})) \quad \text{and} \\
\frac{\partial l}{\partial \zeta} &= \frac{1}{2} \sum_{i=1}^{N} tr(\frac{d\Phi^{-1}}{\partial \zeta}(\Phi - \frac{u_i u_i'}{\sigma_i^2})).
\end{aligned} \tag{46}$$

Proof of theorem 4 - large N, fixed T consistency:

The likelihood equations for σ_i^2, $i = 1, ..., N$, are given by

$$\frac{\partial l_{RE,i}(\theta_0, \sigma_i^2)}{\partial \sigma_i^2} = \frac{T}{2\sigma_i^2} - \frac{1}{2\sigma_i^4} tr(\Phi^{-1} u_i u_i') = 0, \quad i = 1, ..., N. \tag{47}$$

Solving the N equations in (47) for σ_i^2, $i = 1, ..., N$, yields

$$\widetilde{\sigma}_i^2 = \widetilde{\sigma}_i^2(\theta_0) = \frac{1}{T} tr([\Phi(\theta_0)]^{-1} u_i u_i'), \; i = 1, ..., N. \tag{48}$$

After substituting the $\widetilde{\sigma}_i^2(\theta_0) = \widetilde{\sigma}_i^2$ for the σ_i^2 in $\frac{\partial l_{RE}(\theta_0, \sigma_1^2, ..., \sigma_N^2)}{\partial \theta_0} = 0$, we obtain a system of $\dim(\theta_0)$ concentrated likelihood equations: $\frac{\partial l_{RE}(\theta, \widetilde{\sigma}_1^2(\theta_0), ..., \widetilde{\sigma}_N^2(\theta_0))}{\partial \theta}|_{\theta_0} = 0$.

Define L as a $T \times T$ matrix with $L_{s,t} = \rho^{s-t-1}$ if $s > t$ and $L_{s,t} = 0$ if $s \leq t$. Note that $E(\widetilde{\sigma}_i^{-2} \Phi^{-1/2} u_i u_i' \Phi^{-1/2}) = I$. It follows from this and assumption C that $E(\widetilde{\sigma}_i^{-2} y_{i,0} \iota' \Phi^{-1} u_i) = 0$, $E(\widetilde{\sigma}_i^{-2} (y_{i,-1} - \pi y_{i,0}\iota)' \Phi^{-1} u_i) = E(\widetilde{\sigma}_i^{-2} tr(\Phi^{-1} u_i u_i' L')) = tr(L') = 0$ and $E(\Phi - \frac{u_i u_i'}{\widetilde{\sigma}_i^2}) = 0$ and therefore that $E(\frac{\partial l_{RE,i}(\theta_0, \widetilde{\sigma}_i^2)}{\partial \theta_0}) = 0$. Note also that $\sigma_i^{-1} y_i^+$ ($i = 1, ..., N$) are i.i.d. Furthermore note that $\frac{\partial l_{RE,i}(\theta, \sigma_i^2(\widetilde{\theta}))}{\partial \theta}|_{\widetilde{\theta}}$ is continuous in $\widetilde{\theta}$ and $E\left(\sup_{\widetilde{\theta}} \left\| \frac{\partial l_{RE,i}(\theta, \sigma_i^2(\widetilde{\theta}))}{\partial \theta}|_{\widetilde{\theta}} \right\| \right) < \infty$. Hence $\text{plim}_{N \to \infty} \widehat{\theta}_{QMLE} = \theta_0$.

Proof of theorem 4 - large N, T consistency:

We assume $y_{i,0} = 0$, $i = 1, ..., N$ and that θ does not include $\tilde{\pi}$. The extension to the case $y_{i,0} \neq 0$ is straightforward, see the end of the proof. For the extension to a model with time dummies, see Bai (2013, p.294 and S.6 in Supplemental Material).

Let $B(r) = B$ be a $T \times T$ matrix with $B_{i,i} = 1$ and $B_{i+1,i} = -r$ for $i = 1, 2, ..., T-1$, $B_{T,T} = 1$ and $B_{i,j} = 0$ elsewhere. Let $\omega_T = T^{-1}\iota'(\Psi(\zeta))^{-1}\iota$ and $w_T = T^{-1}\iota'(\Psi(z))^{-1}\iota$.

The model is $B(\rho)y_i = \iota\eta_i + \varepsilon_i = u_i$. Let $\Gamma(\rho) = (B(\rho))^{-1}$, then $y_i = \Gamma(\rho)(\iota\eta_i + \varepsilon_i)$.

$$
\begin{aligned}
l_{RE,i}(\theta, s_i^2) &= -\frac{1}{2}\ln|\Phi| - \frac{T}{2}\ln s_i^2 - \frac{1}{2s_i^2}tr(\Phi^{-1}By_iy_i'B') \\
&= -\frac{1}{2}\ln|\Psi| - \frac{1}{2}\ln(1 + T\widetilde{s}_v^2 w_T) - \frac{T}{2}\ln s_i^2 \\
&\quad -\frac{1}{2s_i^2}tr(By_iy_i'B'\Psi^{-1}) + \frac{1}{2s_i^2}\frac{\widetilde{s}_v^2}{1 + T\widetilde{s}_v^2 w_T}(\iota\Psi^{-1}By_iy_i'B'\Psi^{-1}\iota)
\end{aligned}
$$

where $|\Phi| = (1 + T\widetilde{s}_v^2 w_T)|\Psi|$ follows from Lemma 2.1 in Magnus (1982).

Let us define $J_T = -dB/dr$ and $L = J_T\Gamma(r)$ (cf. Bai, 2013, p.300). Note that $By_i = y_i - ry_{i,-1} = u_i + (\rho - r)y_{i,-1} = u_i + (\rho - r)Lu_i = u_i(\theta)$ and

$$
\widetilde{\sigma}_i^2(\theta) = T^{-1}tr([\Phi(\theta)]^{-1}u_i(\theta)u_i'(\theta)) = T^{-1}tr([\Phi(\theta)]^{-1}By_iy_i'B'),\ i = 1, ..., N.
$$

The proof now proceeds as follows. In the first step we replace s_i^2 in the log-likelihood function $\sum_{i=1}^N l_{RE,i}(\theta, s_i^2)$ by $\sigma_i^2 + o_p(1)$ $\forall i \in \{1, ..., N\}$ and prove that $\text{plim}_{N,T\to\infty}\widehat{\theta} = \theta_0$ using a version of Bai's proof. In the second step we note that $\widetilde{\sigma}_i^2(\theta)$, $i = 1, ..., N$, maximize the log-likelihood function for any given θ and we show that $\widetilde{\sigma}_i^2(\widehat{\theta}) = \sigma_i^2 + o_p(1)$ $\forall i \in \{1, ..., N\}$ when $N, T \to \infty$, which justifies replacing s_i^2 in the log-likelihood function $\sum_{i=1}^N l_{RE,i}(\theta, s_i^2)$ by $\sigma_i^2 + o_p(1)$ $\forall i \in \{1, ..., N\}$. The first step:

Let $S_N = N^{-1}\sum_i(y_iy_i'/(\sigma_i^2 + o_p(1)))$ and $\Sigma(\theta_0) = \Gamma(\rho)(\widetilde{\sigma}_v^2\iota\iota' + \Psi(\zeta))\Gamma(\rho)'$. Then

$$
\begin{aligned}
BS_NB'\Phi^{-1} &= B\Sigma(\theta_0)B'\Phi^{-1}(N^{-1}\textstyle\sum_i(\sigma_i^2/(\sigma_i^2 + o_p(1)))) + \\
&\quad B\Gamma(\rho)(N^{-1}\textstyle\sum_i(\sigma_i^2(\eta_i^2\iota\iota'/\sigma_i^2 - \widetilde{\sigma}_v^2\iota\iota')/(\sigma_i^2 + o_p(1))))\Gamma(\rho)'B'\Phi^{-1} + \\
&\quad B\Gamma(\rho)(N^{-1}\textstyle\sum_i(\sigma_i^2(\varepsilon_i\varepsilon_i'/\sigma_i^2 - \Psi(\zeta))/(\sigma_i^2 + o_p(1))))\Gamma(\rho)'B'\Phi^{-1} + \\
&\quad B\Gamma(\rho)(N^{-1}\textstyle\sum_i(\eta_i\iota\varepsilon_i'/(\sigma_i^2 + o_p(1))))\Gamma(\rho)'B'\Phi^{-1} + \\
&\quad B\Gamma(\rho)(N^{-1}\textstyle\sum_i(\eta_i\varepsilon_i\iota'/(\sigma_i^2 + o_p(1))))\Gamma(\rho)'B'\Phi^{-1}.
\end{aligned}
$$

Similarly to the proof of Lemma A.1 in Bai (2013), one can show that as $N, T \to \infty$, uniformly on Θ,

$$T^{-1}tr(B\Gamma(\rho)(N^{-1}\sum_i(\sigma_i^2(\eta_i^2\iota\iota'/\sigma_i^2 - \widetilde{\sigma}_v^2\iota\iota')/(\sigma_i^2 + o_p(1))))\Gamma(\rho)'B'\Phi^{-1}) = o_p(1), \text{ and}$$
$$T^{-2}tr(\iota'\Phi^{-1}B\Gamma(\rho)(N^{-1}\sum_i(\sigma_i^2(\eta_i^2\iota\iota'/\sigma_i^2 - \widetilde{\sigma}_v^2\iota\iota')/(\sigma_i^2 + o_p(1))))\Gamma(\rho)'B'\Phi^{-1}\iota) + o_p(1).$$

Using a lemma similar to Lemma A.1 in Bai, we obtain as $N, T \to \infty$, uniformly on Θ,

$$\begin{aligned} T^{-1}tr(BS_NB'\Phi^{-1}) &= T^{-1}tr(B\Sigma(\theta_0)B'\Phi^{-1}) + o_p(1), \\ T^{-2}tr(\iota'\Phi^{-1}BS_NB'\Phi^{-1}\iota) &= T^{-2}tr(\iota'\Phi^{-1}B\Sigma(\theta_0)B'\Phi^{-1}\iota) + o_p(1), \text{ and} \\ T^{-1}tr(\iota'\Phi^{-1}BS_NB'\Phi^{-1}\iota)/(\iota'\Phi^{-1}\iota) &= T^{-1}tr(\iota'\Phi^{-1}B\Sigma(\theta_0)B'\Phi^{-1}\iota)/(\iota'\Phi^{-1}\iota) + o_p(1). \end{aligned}$$

Furthermore, the diagonal elements of $(\widehat{B}S_N\widehat{B}' - \widehat{B}\Sigma(\theta_0)\widehat{B}')$ are $o_p(1)$ when $N, T \to \infty$, where $\widehat{B} = B(\widehat{\rho})$. Next, using lemmas similar to Lemmas 2 and 3 in Bai, we obtain $\text{plim}_{N,T\to\infty}\widehat{\theta} = \theta_0$.

The second step:

Note that $\Phi(\theta) = \widetilde{s}_v^2\iota\iota' + \Psi$ and $[\Phi(\theta)]^{-1} = \Psi^{-1} - \widetilde{s}_v^2(\widetilde{s}_v^2\iota'\Psi^{-1}\iota + 1)^{-1}\Psi^{-1}\iota\iota'\Psi^{-1}$.

We can easily check the following results, cf. Bai:

$\frac{1}{T}tr([\Phi(\theta)]^{-1}Lu_iu_i'L') =:$
$\frac{1}{T}tr([\Phi(\theta)]^{-1}L(\varepsilon_i\varepsilon_i' - \sigma_i^2\Psi_0)L') =$
$W_i = \varepsilon_i\varepsilon_i' - \sigma_i^2\Psi_0$
$tr(\Psi^{-1}LW_iL') = \sum_{t=1}^{T-1} l_{t+1}^{-2}\sum_{h=1}^{t}\rho^{t-h}\sum_{k=1}^{t}\rho^{t-k}W_{i,hk}$
$\sup_\Psi\left|\frac{1}{T}tr(\Psi^{-1}L(\varepsilon_i\varepsilon_i' - \sigma_i^2\Psi_0)L')\right| = O_p(1).$
$\sup_\Psi\left|\frac{1}{T^2}tr(\Psi^{-1}\iota\iota'\Psi^{-1}L(\varepsilon_i\varepsilon_i' - \sigma_i^2\Psi_0)L')\right| = O_p(1).$
$\frac{1}{T}tr([\Phi(\theta)]^{-1}L\iota\eta_i\varepsilon_i'L') =$
$\sup_\Psi\left|\frac{1}{T}tr(\Psi^{-1}L\iota\eta_i\varepsilon_i'L')\right| = O_p(1).$
$\sup_\Psi\left|\frac{1}{T^2}tr(\Psi^{-1}\iota\iota'\Psi^{-1}L\iota\eta_i\varepsilon_i'L')\right| = O_p(1).$
$\frac{1}{T}tr([\Phi(\theta)]^{-1}L\iota\iota'L'\eta_i^2) =$
$\sup_\Psi\left|\frac{1}{T}tr(\Psi^{-1}L\iota\iota'L'\eta_i^2)\right| = O_p(1).$
$\sup_\Psi\left|\frac{1}{T^2}tr(\Psi^{-1}\iota\iota'\Psi^{-1}L\iota\iota'L'\eta_i^2)\right| = O_p(1).$
$\frac{1}{T}tr([\Phi(\theta)]^{-1}Lu_iu_i') =:$
$\frac{1}{T}tr([\Phi(\theta)]^{-1}L(\varepsilon_i\varepsilon_i' - \sigma_i^2\Psi_0)) =$

$tr(\Psi^{-1}LW_i) = \sum_{k=0}^{T-2} \rho^k \sum_{t=k+2}^{T} l_t^{-2} \varepsilon_{i,t-k-1}\varepsilon_{i,t}$
$\sup_\Psi \left|\frac{1}{T} tr(\Psi^{-1}L(\varepsilon_i\varepsilon_i' - \sigma_i^2\Psi_0))\right| = O_p(1).$
$\sup_\Psi \left|\frac{1}{T^2} tr(\Psi^{-1}\iota\iota'\Psi^{-1}L(\varepsilon_i\varepsilon_i' - \sigma_i^2\Psi_0))\right| = O_p(1).$
$\frac{1}{T} tr([\Phi(\theta)]^{-1}L\iota\eta_i\varepsilon_i') =$
$\sup_\Psi \left|\frac{1}{T} tr(\Psi^{-1}L\iota\eta_i\varepsilon_i')\right| = O_p(1).$
$\sup_\Psi \left|\frac{1}{T^2} tr(\Psi^{-1}\iota\iota'\Psi^{-1}L\iota\eta_i\varepsilon_i')\right| = O_p(1).$
$\frac{1}{T} tr([\Phi(\theta)]^{-1}L\iota\iota'\eta_i^2) =$
$\sup_\Psi \left|\frac{1}{T} tr(\Psi^{-1}L\iota\iota'\eta_i^2)\right| = O_p(1).$
$\sup_\Psi \left|\frac{1}{T^2} tr(\Psi^{-1}\iota\iota'\Psi^{-1}L\iota\iota'\eta_i^2)\right| = O_p(1).$
$\frac{1}{T}\sigma_i^2 tr([\Phi(\theta)]^{-1}L\Psi_0 L') =$
$\sup_\Psi \left|\frac{1}{T}\sigma_i^2 tr(\Psi^{-1}L\Psi_0 L')\right| = O(1).$
$\sup_\Psi \left|\frac{1}{T^2}\sigma_i^2 tr(\Psi^{-1}\iota\iota'\Psi^{-1}L\Psi_0 L')\right| = O(T^{-1}).$
$\frac{1}{T}\sigma_i^2 tr([\Phi(\theta)]^{-1}L\Psi_0) =$
$\sup_\Psi \frac{1}{T}\sigma_i^2 tr(\Psi^{-1}L\Psi_0) = 0.$
$\sup_\Psi \left|\frac{1}{T^2}\sigma_i^2 tr(\Psi^{-1}\iota\iota'\Psi^{-1}L\Psi_0)\right| = O(T^{-1}).$

Now consider $T^{-1}tr([\Phi(\widehat{\theta})]^{-1}u_iu_i')$. Note that

$$T^{-1}tr([\Phi(\theta)]^{-1}u_iu_i') = T^{-1}u_i'\Psi^{-1}u_i - T^{-1}\widetilde{s}_v^2(\widetilde{s}_v^2\iota'\Psi^{-1}\iota + 1)^{-1}u_i'\Psi^{-1}\iota\iota'\Psi^{-1}u_i.$$

$T^{-1}\varepsilon_i'([\Psi(\widehat{\theta})]^{-1})\varepsilon_i =:$
$T^{-1}tr([\Psi(\theta)]^{-1}(\varepsilon_i\varepsilon_i' - \sigma_i^2\Psi_0)) = T^{-1}\sum_{t=1}^{T}(\varepsilon_{i,t}^2 - \sigma_i^2\lambda_t^2)/l_t^2 =$
$T^{-1}\sum_{t=1}^{T}((\varepsilon_{i,t}^2/\lambda_t^2 - \sigma_i^2)(\lambda_t^2/l_t^2)) = O_p(T^{-1/2})$ uniformly on Θ.
Hence $T^{-1}tr([\Psi(\widehat{\theta})]^{-1}(\varepsilon_i\varepsilon_i' - \sigma_i^2\Psi_0)) = O_p(T^{-1/2})$
$T^{-1}\sigma_i^2 tr([\Psi(\widehat{\theta})]^{-1}\Psi_0) = \sigma_i^2 + O_p(T^{-1/2}N^{-1/2})$ because $\widehat{\lambda}_t^2 - \lambda_t^2 = O_p(N^{-1/2}) + O_p(T^{-1})$.
Hence $T^{-1}\varepsilon_i'([\Psi(\widehat{\theta})]^{-1})\varepsilon_i = \sigma_i^2 + O_p(T^{-1/2}) + O_p(T^{-1/2}N^{-1/2})$
Similarly $T^{-1}\iota'([\Psi(\widehat{\theta})]^{-1})\varepsilon_i = O_p(T^{-1/2})$
$T^{-1}\iota'[\Phi(\theta)]^{-1}\iota\eta_i^2 = T^{-1}\iota\Psi^{-1}\iota\eta_i^2 - T^{-1}\widetilde{s}_v^2(\widetilde{s}_v^2\iota'\Psi^{-1}\iota + 1)^{-1}(\iota\Psi^{-1}\iota)^2\eta_i^2 =$
$(\widetilde{s}_v^2\iota'\Psi^{-1}\iota + 1)^{-1}(T^{-1}\iota\Psi^{-1}\iota\eta_i^2)$
$(\widetilde{s}_v^2\iota'\Psi^{-1}\iota + 1)^{-1}(\iota\Psi^{-1}\iota) - 1/\widetilde{s}_v^2 = o(1)$ uniformly on Θ.
Hence $T^{-1}\iota'[\Phi(\widehat{\theta})]^{-1}\iota\eta_i^2 = O_p(T^{-1})$
We also have $T^{-1}\iota'([\Psi(\widehat{\theta})]^{-1})\varepsilon_i\eta_i = O_p(T^{-1/2})$
We conclude $T^{-1}tr([\Phi(\widehat{\theta})]^{-1}u_iu_i') = \sigma_i^2 + o_p(1)$

Since $\widehat{\rho} = \rho + o_p(1)$ we conclude

$$\begin{aligned}
\widetilde{\sigma}_i^2(\widehat{\theta}) &= T^{-1}tr([\Phi(\widehat{\theta})]^{-1}By_iy_i'B') \\
&= T^{-1}tr([\Phi(\widehat{\theta})]^{-1}(u_i + (\rho - \widehat{\rho})Lu_i)(u_i + (\rho - \widehat{\rho})Lu_i)') \\
&= T^{-1}tr([\Phi(\widehat{\theta})]^{-1}u_iu_i') + o_p(1) \\
&= \sigma_i^2 + o_p(1) \ \forall i \text{ when } N, T \to \infty. \qquad \square
\end{aligned}$$

The extension to the case $y_{i,0} \neq 0$ is straightforward: we only need to redefine S_N as $S_N = N^{-1}\sum_i((y_i - \pi y_{i,0})(y_i - \pi y_{i,0})'/(\sigma_i^2 + o_p(1)))$. The vector θ, which now also includes π, is again identified and $\widehat{\theta}$ is again joint N, T consistent (and asymptotically normal).

Note that in the case $y_{i,0} \neq 0$, the REQMLE discussed in Bai (2013, S.3 in Supplemental Material) is different from our REQMLE. Our REQMLE is equal to Chamberlain's REMLE which uses the decomposition of the 'correlated effects' η_i given in (28). Instead of using (28), Bai adds an equation for the initial condition to the model for y_i, i.e., $y_{i,0} = \delta_0 + \phi\eta_i + \varepsilon_{i,0}$, where δ_0 and ϕ are parameters and $\varepsilon_{i,0}$ is a random variable with $E(\varepsilon_{i,0}) = 0$ and $Var(\varepsilon_{i,0}) = \sigma_0^2$, and then derives and estimates a factor model for $y_i^+ = (y_{i,0},\, y_i')' : y_i^+ = \delta^+ + \Gamma^+(\phi, \iota')'\eta_i + \Gamma^+\varepsilon_i^+$, where δ^+ is a parameter vector and Γ^+ has exactly the same form as Γ but dimension $(T+1) \times (T+1)$ and $\varepsilon_i^+ = (\varepsilon_{i,0},\, \varepsilon_i')'$. Note that Bai's REQMLE does not estimate π, whereas our REQMLE does not estimate ϕ.

Proof of theorem 5:

The proof is similar to the proof of theorem 4 for the case $T \to \infty$ with the following changes: replace θ_0 by $\overline{\theta}_0$ and σ_i^2 by $\overline{\sigma}_i^2$, $i = 1, ..., N$. Then in the first step we replace s_i^2 in the log-likelihood function $\sum_{i=1}^N l_{RE,i}(\theta, s_i^2)$ by $\overline{\sigma}_i^2 + o_p(1)$ $\forall i \in \{1, ..., N\}$ and prove that $\text{plim}_{N,T\to\infty}\widehat{\theta} = \overline{\theta}_0$ using a version of Bai's (2013) proof. To prove lemmas similar to Lemmas A.1, 2 and 3 in Bai, we make use of inter alia assumption C$'$(iii), which implies that $\text{plim}_{N,T\to\infty}N^{-1}\sum_{i=1}^N \left(diag(\sigma_{i,1}^2, ..., \sigma_{i,T}^2)/(\overline{\sigma}_i^2 + o_p(1))\right) = \Psi(\overline{\theta}_0)$ and that $\text{plim}_{N,T\to\infty}N^{-1}\sum_{i=1}^N(\eta_i^2/(\overline{\sigma}_i^2 + o_p(1))) = \widetilde{\sigma}_v^2$. The full proof of step 1 is given below. The second step is straightforward: $\text{plim}_{N,T\to\infty}\widetilde{\sigma}_i^2(\widehat{\theta}) = \text{plim}_{T\to\infty}\widetilde{\sigma}_i^2(\overline{\theta}_0) = \overline{\sigma}_i^2 \neq 0$, $i = 1, ..., N$, by assumption C$'$(iii).

Proof of step 1:

Let $S_N = N^{-1}\sum_i(y_iy_i'/(\overline{\sigma}_i^2 + o_p(1)))$, let $\Sigma_N(\theta_0) = \Gamma(\rho)(\widetilde{\sigma}_{v,N}^2\iota\iota' + \Psi(\zeta_N))\Gamma(\rho)'$ with

$\widetilde{\sigma}^2_{v,N} = N^{-1}\sum_{i=1}^{N}\left(\eta_i^2/(\overline{\sigma}_i^2 + o_p(1))\right)$ and $\Psi(\zeta_N) = N^{-1}\sum_{i=1}^{N}(diag(\overline{\sigma}_i^2 + o_p(1), \sigma_{i,2}^2, ..., \sigma_{i,T}^2)/(\overline{\sigma}_i^2 + o_p(1)))$, and let $\widetilde{\Psi}_N(\theta_0) = N^{-1}\sum_{i=1}^{N}\left(diag(\sigma_{i,1}^2, ..., \sigma_{i,T}^2)/(\overline{\sigma}_i^2 + o_p(1))\right)$. Then

$$\begin{aligned} BS_N B'\Phi^{-1} &= B\Sigma_N(\theta_0)B'\Phi^{-1} + \\ &\quad B\Gamma(\rho)(\widetilde{\Psi}_N(\theta_0) - \Psi(\zeta_N))\Gamma(\rho)'B'\Phi^{-1} + \\ &\quad B\Gamma(\rho)(N^{-1}\textstyle\sum_i(\varepsilon_i\varepsilon_i'/(\overline{\sigma}_i^2 + o_p(1))) - \widetilde{\Psi}_N(\theta_0))\Gamma(\rho)'B'\Phi^{-1} + \\ &\quad B\Gamma(\rho)(N^{-1}\textstyle\sum_i(\eta_i\iota\varepsilon_i'/(\overline{\sigma}_i^2 + o_p(1))))\Gamma(\rho)'B'\Phi^{-1} + \\ &\quad B\Gamma(\rho)(N^{-1}\textstyle\sum_i(\eta_i\varepsilon_i\iota'/(\overline{\sigma}_i^2 + o_p(1))))\Gamma(\rho)'B'\Phi^{-1}. \end{aligned}$$

Using assumption C′(iii), which implies that $\text{plim}_{N,T\to\infty}N^{-1}\sum_{i=1}^{N}(\sigma_{i,1}^2/(\overline{\sigma}_i^2 + o_p(1))) = 1$, and $B\Gamma(\rho) = I + (\rho - r)L^0$, where L^0 denotes L at $r = \rho$, it is easily seen that as $N, T \to \infty$, uniformly on Θ,

$$\begin{aligned} T^{-1}tr(B\Gamma(\rho)(\widetilde{\Psi}_N(\theta_0) - \Psi(\zeta_N))\Gamma(\rho)'B'\Phi^{-1}) &= o_p(1), \\ T^{-2}tr(\iota'\Phi^{-1}B(\widetilde{\Psi}_N(\theta_0) - \Psi(\zeta_N))B'\Phi^{-1}\iota) &= o_p(1), \\ T^{-1}tr(\iota'\Phi^{-1}B(\widetilde{\Psi}_N(\theta_0) - \Psi(\zeta_N))B'\Phi^{-1}\iota)/(\iota'\Phi^{-1}\iota) &= o_p(1). \end{aligned}$$

Using these results, assumption C′ and a lemma similar to Lemma A.1 in Bai (2013), we also obtain that as $N, T \to \infty$, uniformly on Θ,

$$\begin{aligned} T^{-1}tr(BS_NB'\Phi^{-1}) &= T^{-1}tr(B\Sigma_N(\theta_0)B'\Phi^{-1}) + o_p(1), \\ T^{-2}tr(\iota'\Phi^{-1}BS_NB'\Phi^{-1}\iota) &= T^{-2}tr(\iota'\Phi^{-1}B\Sigma_N(\theta_0)B'\Phi^{-1}\iota) + o_p(1), \text{ and} \\ T^{-1}tr(\iota'\Phi^{-1}BS_NB'\Phi^{-1}\iota)/(\iota'\Phi^{-1}\iota) &= T^{-1}tr(\iota'\Phi^{-1}B\Sigma_N(\theta_0)B'\Phi^{-1}\iota)/(\iota'\Phi^{-1}\iota) + o_p(1). \end{aligned}$$

Furthermore, the diagonal elements of $(\widehat{B}S_N\widehat{B}' - \widehat{B}\Sigma_N(\theta_0)\widehat{B}')$ are $o_p(1)$ when $N, T \to \infty$, where $\widehat{B} = B(\widehat{\rho})$. Finally, using lemmas similar to Lemmas 2 and 3 in Bai, we obtain that $\text{plim}_{N,T\to\infty}\widehat{\theta} = \theta_0$. □

Proof of theorem 7:

The derivation of the asymptotic distribution of $\widehat{\theta} = \widehat{\theta}_{REQMLE}$ is largely similar to the derivation in Bai (2013) but in our case we also have to consider some potential bias terms that are related to the presence of the $\widetilde{\sigma}_i^2(\widehat{\theta})$:

In the case of the distribution of $\widehat{\rho} = \widehat{\rho}_{REQMLE}$, we need to consider the following additional terms related to the $\widetilde{\sigma}_i^2(\widehat{\theta})$:

$(NT)^{-1/2}\sum_{i=1}^N\left(\frac{\varepsilon_i'\Psi^{-1}L\iota\eta_i}{\widetilde{\sigma}_i^2(\widehat{\theta})}-\frac{\varepsilon_i'\Psi^{-1}L\iota\eta_i}{\sigma_i^2}\right)-\frac{\iota'\widehat{\Psi}^{-1}L\iota}{T\widehat{\omega}}(NT)^{-1/2}\sum_{i=1}^N\left(\frac{\varepsilon_i'\Psi^{-1}\iota\eta_i}{\widetilde{\sigma}_i^2(\widehat{\theta})}-\frac{\varepsilon_i'\Psi^{-1}\iota\eta_i}{\sigma_i^2}\right).$

Recall from Bai that $\frac{\iota'\widehat{\Psi}^{-1}L\iota}{T\widehat{\omega}_T}=(1-\rho)^{-1}+O_p(1/T)$.

Next consider $-(NT)^{-1/2}\sum_{i=1}^N\left(\frac{\varepsilon_i'\Psi^{-1}\iota\eta_i}{\widetilde{\sigma}_i^2(\widehat{\theta})}-\frac{\varepsilon_i'\Psi^{-1}\iota\eta_i}{\sigma_i^2}\right)=(NT)^{-1/2}\sum_{i=1}^N\frac{\varepsilon_i'\Psi^{-1}\iota\eta_i}{\sigma_i^4}(\widetilde{\sigma}_i^2(\widehat{\theta})-\sigma_i^2)+o_p(1)=(NT)^{-1/2}\sum_{i=1}^N\sum_{t=1}^T\frac{\varepsilon_{i,t}\eta_i}{\lambda_t^2\sigma_i^4}(T^{-1}\sum_{s=1}^T(\frac{\varepsilon_{i,s}^2}{\lambda_s^2}-\sigma_i^2))+o_p(1)$.

This factor has expected value: $(N/T)^{1/2}(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T\frac{E(\varepsilon_{i,t}^3)\eta_i}{\lambda_t^4\sigma_i^4}$.

The associated potentially nonnegligible bias term is the product of the preceding expression and $(1-\rho)^{-1}$, cf. top of p. 312 in Bai. The other term we need to consider is:

$(NT)^{-1/2}\sum_{i=1}^N\left(\frac{\varepsilon_i'\Psi^{-1}L\iota\eta_i}{\widetilde{\sigma}_i^2(\widehat{\theta})}-\frac{\varepsilon_i'\Psi^{-1}L\iota\eta_i}{\sigma_i^2}\right)=-(NT)^{-1/2}\sum_{i=1}^N\frac{\varepsilon_i'\Psi^{-1}L\iota\eta_i}{\sigma_i^4}(\widetilde{\sigma}_i^2(\widehat{\theta})-\sigma_i^2)+o_p(1)$.

This term has expected value: $-(N/T)^{1/2}(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T\frac{E(\varepsilon_{i,t}^3)m_t\eta_i}{\lambda_t^4\sigma_i^4}$ where $m_t=1+\rho+...+\rho^{t-2}\to(1-\rho)^{-1}$ so that $(N/T)^{1/2}(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T\frac{E(\varepsilon_{i,t}^3)m_t\eta_i}{\lambda_t^4\sigma_i^4}\to(1-\rho)^{-1}(N/T)^{1/2}(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T\frac{E(\varepsilon_{i,t}^3)\eta_i}{\lambda_t^4\sigma_i^4}$.

Thus the additional bias terms cancel out.

In the case of the distribution of $\widehat{\widetilde{\sigma}}_v^2=\widehat{\widetilde{\sigma}}_{v,REQMLE}^2$, there is an additional bias term that is related to the $\widetilde{\sigma}_i^2(\widehat{\theta})$:

$(NT)^{-1}\sum_{i=1}^N\left(\frac{\varepsilon_i'\Psi^{-1}\iota\eta_i}{\widetilde{\sigma}_i^2(\widehat{\theta})}-\frac{\varepsilon_i'\Psi^{-1}\iota\eta_i}{\sigma_i^2}\right)=-(NT)^{-1}\sum_{i=1}^N\frac{\varepsilon_i'\Psi^{-1}\iota\eta_i}{\sigma_i^4}(\widetilde{\sigma}_i^2(\widehat{\theta})-\sigma_i^2)+o_p(1)=-(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T\frac{\varepsilon_{i,t}\eta_i}{\lambda_t^2\sigma_i^4}(T^{-1}\sum_{s=1}^T(\frac{\varepsilon_{i,s}^2}{\lambda_s^2}-\sigma_i^2))+o_p(1)$.

The additional bias term is: $-(2/(T\omega_T))(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T\frac{E(\varepsilon_{i,t}^3)\eta_i}{\lambda_t^4\sigma_i^4}=\frac{b}{T}=O(1/T)$.

In the case of the asymptotic distributions of $\widehat{\lambda}_t^2=\widehat{\lambda}_{t,REQMLE}^2$, $t=1,...,T$, we need to consider some additional terms related to $\widetilde{\sigma}_i^2(\widehat{\theta})$:

First consider the term $N^{-1}\sum_{i=1}^N\frac{\varepsilon_i\varepsilon_i'}{\sigma_i^4}(\widetilde{\sigma}_i^2(\widehat{\theta})-\sigma_i^2)$:

$N^{-1}\sum_{i=1}^N\frac{\varepsilon_{i,t}^2}{\sigma_i^4}(\widetilde{\sigma}_i^2(\widehat{\theta})-\sigma_i^2)=N^{-1}\sum_{i=1}^N\frac{\varepsilon_{i,t}^2}{\sigma_i^4}(T^{-1}\sum_{s=1}^T(\frac{\varepsilon_{i,s}^2}{\lambda_s^2}-\sigma_i^2))+o_p(1)=$
$(NT)^{-1}\sum_{i=1}^N\frac{\varepsilon_{i,t}^4-\sigma_i^2\lambda_t^2\varepsilon_{i,t}^2}{\sigma_i^4\lambda_t^2}+(NT)^{-1}\sum_{i=1}^N\sum_{s\neq t}^T\frac{(\varepsilon_{i,s}^2-\sigma_i^2\lambda_s^2)\varepsilon_{i,t}^2}{\sigma_i^4\lambda_s^2}+o_p(1)$.

This term has expected value: $(NT)^{-1}\sum_{i=1}^N\frac{E(\varepsilon_{i,t}^4)-\sigma_i^2\lambda_t^2E(\varepsilon_{i,t}^2)}{\sigma_i^4\lambda_t^2}=O(1/T)$.

Next consider $(NT)^{-1}\sum_{i=1}^N\frac{\varepsilon_i'\Psi^{-1}\iota\eta_i}{\sigma_i^4}(\widetilde{\sigma}_i^2(\widehat{\theta})-\sigma_i^2)=(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T\frac{\varepsilon_{i,t}\eta_i}{\lambda_t^2\sigma_i^4}(T^{-1}\sum_{s=1}^T(\frac{\varepsilon_{i,s}^2}{\lambda_s^2}-\sigma_i^2))+o_p(1)$.

This term has expected value: $(1/T)(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{E(\varepsilon_{i,t}^{3})\eta_i}{\lambda_t^4\sigma_i^4}=O(1/T)$.

Finally, consider the term $(NT)^{-1}\sum_{i=1}^{N}\frac{\varepsilon_i\iota'}{\sigma_i^4}(\varepsilon_i'\Psi^{-1}\iota)(\widetilde{\sigma}_i^2(\widehat{\theta})-\sigma_i^2)$:

$(NT)^{-1}\sum_{i=1}^{N}\frac{\varepsilon_{i,t}}{\sigma_i^4}(\varepsilon_i'\Psi^{-1}\iota)(\widetilde{\sigma}_i^2(\widehat{\theta})-\sigma_i^2)=$

$(NT)^{-1}\sum_{i=1}^{N}\frac{\varepsilon_{i,t}}{\sigma_i^4}(\varepsilon_i'\Psi^{-1}\iota)(T^{-1}\sum_{s=1}^{T}(\frac{\varepsilon_{i,s}^2}{\lambda_s^2}-\sigma_i^2))+o_p(1)=$

$(NT)^{-1}\sum_{i=1}^{N}\frac{\varepsilon_{i,t}^2}{\sigma_i^4\lambda_t^2}(T^{-1}\sum_{s=1}^{T}(\frac{\varepsilon_{i,s}^2}{\lambda_s^2}-\sigma_i^2))+o_p(1)$.

This term has expected value: $T^{-1}(NT)^{-1}\sum_{i=1}^{N}\frac{E(\varepsilon_{i,t}^4)-\sigma_i^2\lambda_t^2E(\varepsilon_{i,t}^2)}{\sigma_i^4\lambda_t^4}=O(1/T^2)$.

As the expected values of the three preceding terms are $O(1/T)$ or $O(1/T^2)$, they do not affect the asymptotic distributions of $\widehat{\lambda}_t^2=\widehat{\lambda}_{t,REQMLE}^2$, $t=1,...,T$, as long as $\sqrt{N}/T\to 0$.

To prove (33), we use results from the second step in the proof of theorem 4.

They imply $\sqrt{T}(\widetilde{\sigma}_i^2(\widehat{\theta})-\sigma_i^2)=\sqrt{T}(T^{-1}tr([\Psi(\widehat{\theta})]^{-1}\varepsilon_i\varepsilon_i')-\sigma_i^2)+O_p(T^{-1/2})+O_p(N^{-1/2})$, $i=1,...,N$. Note that

$T^{-1/2}(tr(([\Psi(\widehat{\theta})]^{-1}-[\Psi(\theta_0)]^{-1})\varepsilon_i\varepsilon_i'))=-T^{-1/2}\sum_{t=1}^{T}(\frac{\varepsilon_{i,t}^2}{\lambda_t^4}(\widehat{\lambda}_t^2-\lambda_t^2+o_p(1)))=$

$-(NT)^{-1/2}\sum_{t=1}^{T}(\frac{\varepsilon_{i,t}^2}{\lambda_t^4}N^{1/2}(\widehat{\lambda}_t^2-\lambda_t^2+o_p(1)))=$

$-T^{-1/2}N^{-1}\sum_{t=1}^{T}(\frac{\varepsilon_{i,t}^4}{\lambda_t^4\sigma_i^2})=-T^{1/2}N^{-1}T^{-1}\sum_{t=1}^{T}(\frac{\varepsilon_{i,t}^4}{\lambda_t^4\sigma_i^2})+O_p(N^{-1/2})$.

Hence $\sqrt{T}(\widetilde{\sigma}_i^2(\widehat{\theta})-\sigma_i^2)=\sqrt{T}(T^{-1}tr([\Psi(\theta_0)]^{-1}\varepsilon_i\varepsilon_i')-\sigma_i^2)+o_p(1)$, $i=1,...,N$.

The proof of (34) is similar to that of Theorem 1 in Bai and makes use of the preceding results.

The proofs of the remaining claims are straightforward.

Proof of theorem 8:

The proof is straightforward: to prove theorem 8, one can follow e.g. White (1994).

Large N, fixed T limiting distribution:

Let $\widetilde{\Delta} y_i = y_i - y_{i,0}\iota$, $\widetilde{\Delta} y_{i,-1} = y_{i,-1} - y_{i,0}\iota$ and $u_i = \widetilde{\Delta} y_i - \rho \widetilde{\Delta} y_{i,-1}$. Then the log-likelihood function for the non-stationary panel AR(1) model with fixed effects is given by $l = l_{FE} = \sum_{i=1}^N l_{FE,i}(\theta_0, \sigma_i^2)$ with $\theta_0 = (\rho, \widetilde{\sigma}_v^2, \zeta')'$ and

$$\begin{aligned} l_{FE,i}(\theta_0, \sigma_i^2) &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\Phi| - \frac{T}{2}\ln\sigma_i^2 - \\ &\quad \frac{1}{2\sigma_i^2}(\widetilde{\Delta} y_i - \rho\widetilde{\Delta} y_{i,-1})'\Phi^{-1}(\widetilde{\Delta} y_i - \rho\widetilde{\Delta} y_{i,-1}). \end{aligned}$$

Assume that $\Psi = I$ so that $\Phi = \widetilde{\sigma}_v^2 \iota\iota' + I$. Then $\Phi^{-1} = Q + \frac{1}{1+T\widetilde{\sigma}_v^2}\frac{1}{T}\iota\iota'$ and

$$\begin{aligned} l_{FE,i}(\theta_0, \sigma_i^2) &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln(1+T\widetilde{\sigma}_v^2) - \frac{T}{2}\ln\sigma_i^2 - \\ &\quad \frac{1}{2\sigma_i^2}(\widetilde{\Delta} y_i - \rho\widetilde{\Delta} y_{i,-1})'Q(\widetilde{\Delta} y_i - \rho\widetilde{\Delta} y_{i,-1}) - \\ &\quad \frac{1}{2\sigma_i^2}\frac{1}{1+T\widetilde{\sigma}_v^2}\frac{1}{T}(\iota'(\widetilde{\Delta} y_i - \rho\widetilde{\Delta} y_{i,-1}))^2. \end{aligned}$$

$$\begin{aligned} \frac{\partial l}{\partial \rho} &= \sum_{i=1}^N (\frac{1}{\sigma_i^2} u_i'\Phi^{-1}\widetilde{\Delta} y_{i,-1}), \\ \frac{\partial l}{\partial \widetilde{\sigma}_v^2} &= -\frac{NT}{2(1+T\widetilde{\sigma}_v^2)} + \frac{1}{2(1+T\widetilde{\sigma}_v^2)^2}\sum_{i=1}^N \frac{(\iota' u_i)^2}{\sigma_i^2}. \end{aligned}$$

Suppose the σ_i^2 are known:

$$\begin{aligned} \frac{\partial^2 l(\theta_0, \sigma_i^2)}{\partial \rho^2} &= -\sum_{i=1}^N (\frac{1}{\sigma_i^2}(\widetilde{\Delta} y_{i,-1})'\Phi^{-1}(\widetilde{\Delta} y_{i,-1})), \\ \frac{\partial^2 l(\theta_0, \sigma_i^2)}{\partial \rho \partial \widetilde{\sigma}_v^2} &= -\frac{1}{(1+T\widetilde{\sigma}_v^2)^2}\sum_{i=1}^N \frac{(\iota' u_i)(\iota'(\widetilde{\Delta} y_{i,-1}))}{\sigma_i^2}, \\ \frac{\partial^2 l(\theta_0, \sigma_i^2)}{\partial (\widetilde{\sigma}_v^2)^2} &= \frac{NT^2}{2(1+T\widetilde{\sigma}_v^2)^2} - \frac{T}{(1+T\widetilde{\sigma}_v^2)^3}\sum_{i=1}^N \frac{(\iota' u_i)^2}{\sigma_i^2}. \end{aligned}$$

$$plim_{N\to\infty}\frac{1}{N}\frac{\partial l(\theta_0)}{\partial\rho}\frac{\partial l(\theta_0)}{\partial\rho} = plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{\sigma_i^4}(\widetilde{\Delta}y_{i,-1})'\Phi^{-1}u_iu_i'\Phi^{-1}(\widetilde{\Delta}y_{i,-1})),$$

$$plim_{N\to\infty}\frac{1}{N}\frac{\partial l(\theta_0)}{\partial\rho}\frac{\partial l(\theta_0)}{\partial\widetilde{\sigma}_v^2} =$$

$$plim_{N\to\infty}\frac{1}{2(1+T\widetilde{\sigma}_v^2)^2}\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{\sigma_i^4}(\widetilde{\Delta}y_{i,-1})'\Phi^{-1}u_i(\iota'u_i)^2),$$

$$plim_{N\to\infty}\frac{1}{N}\frac{\partial l(\theta_0)}{\partial\widetilde{\sigma}_v^2}\frac{\partial l(\theta_0)}{\partial\widetilde{\sigma}_v^2} =$$

$$plim_{N\to\infty}(-\frac{T^2}{4(1+T\widetilde{\sigma}_v^2)^2}+\frac{1}{4(1+T\widetilde{\sigma}_v^2)^4}\frac{1}{N}\sum_{i=1}^{N}\frac{(\iota'u_i)^4}{\sigma_i^4}).$$

Suppose the σ_i^2 are unknown: substitute $\widetilde{\sigma}_i^2 = \widetilde{\sigma}_i^2(\theta_0)$ for σ_i^2 in $\frac{\partial l(\theta_0)}{\partial\theta_0}\frac{\partial l(\theta_0)}{\partial\theta_0'}$,

$$\begin{aligned}\frac{\partial\widetilde{\sigma}_i^2}{\partial\rho} &= -\frac{2}{T}(\widetilde{\Delta}y_{i,-1})'\Phi^{-1}u_i,\\ \frac{\partial\widetilde{\sigma}_i^2}{\partial\widetilde{\sigma}_v^2} &= -\frac{1}{T(1+T\widetilde{\sigma}_v^2)^2}(\iota'u_i)^2.\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l(\theta_0)}{\partial\rho^2} &= \frac{\partial^2 l(\theta_0,\widetilde{\sigma}_i^2)}{\partial\rho^2} - \sum_{i=1}^{N}(\frac{1}{\widetilde{\sigma}_i^4}(\widetilde{\Delta}y_{i,-1})'\Phi^{-1}u_i\frac{\partial\widetilde{\sigma}_i^2}{\partial\rho}),\\ \frac{\partial^2 l(\theta_0)}{\partial\rho\partial\widetilde{\sigma}_v^2} &= \frac{\partial^2 l(\theta_0,\widetilde{\sigma}_i^2)}{\partial\rho\partial\widetilde{\sigma}_v^2} - \frac{1}{2(1+T\widetilde{\sigma}_v^2)^2}\sum_{i=1}^{N}\left(\frac{(\iota'u_i)^2}{\widetilde{\sigma}_i^4}\frac{\partial\widetilde{\sigma}_i^2}{\partial\rho}\right),\\ \frac{\partial^2 l(\theta_0)}{\partial(\widetilde{\sigma}_v^2)^2} &= \frac{\partial^2 l(\theta_0,\widetilde{\sigma}_i^2)}{\partial(\widetilde{\sigma}_v^2)^2} - \frac{1}{2(1+T\widetilde{\sigma}_v^2)^2}\sum_{i=1}^{N}\left(\frac{(\iota'u_i)^2}{\widetilde{\sigma}_i^4}\frac{\partial\widetilde{\sigma}_i^2}{\partial\widetilde{\sigma}_v^2}\right).\end{aligned}$$

General case: $\Phi = \Phi(\xi) = \widetilde{\sigma}_v^2 \iota\iota' + \Psi(\zeta)$ where $\xi = (\widetilde{\sigma}_v^2\ \zeta')'$. Reformulate the RE model as $\widetilde{\Delta} y_{i,-1} = \rho \widetilde{\Delta} y_{i,-1} + \breve{\pi} \iota y_{i,0} + u_i$ with $\breve{\pi} = (\pi - 1)(1 - \rho)$. In the FE model $\breve{\pi} = 0$. We derive components of the limiting variance of the REMLE for $(\rho\ \breve{\pi}\ \xi')'$.

$$
\begin{aligned}
\frac{\partial l}{\partial \rho} &= \sum_{i=1}^{N} (\frac{1}{\sigma_i^2} u_i' \Phi^{-1} \widetilde{\Delta} y_{i,-1}), \qquad \frac{\partial l}{\partial \breve{\pi}} = \sum_{i=1}^{N} (\frac{1}{\sigma_i^2} u_i' \Phi^{-1} \iota y_{i,0}), \\
\frac{\partial l}{\partial \xi} &= \sum_{i=1}^{N} \left(\frac{1}{2\sigma_i^2} (\frac{\partial vec \Phi^{-1}}{\partial \xi'})' vec(\sigma_i^2 \Phi - u_i u_i') \right) \\
&= -\sum_{i=1}^{N} \left(\frac{1}{2\sigma_i^2} (\frac{\partial vec \Phi}{\partial \xi'})' \Phi^{-1} \otimes \Phi^{-1} vec(\sigma_i^2 \Phi - u_i u_i') \right).
\end{aligned}
$$

Suppose the σ_i^2 are known:

$$
\begin{aligned}
\frac{\partial^2 l(\theta_0, \sigma_i^2)}{\partial \rho^2} &= -\sum_{i=1}^{N} (\frac{1}{\sigma_i^2} (\widetilde{\Delta} y_{i,-1})' \Phi^{-1} (\widetilde{\Delta} y_{i,-1})), \\
\frac{\partial^2 l(\theta_0, \sigma_i^2)}{\partial \rho \partial \breve{\pi}} &= -\sum_{i=1}^{N} (\frac{1}{\sigma_i^2} (\widetilde{\Delta} y_{i,-1})' \Phi^{-1} (\iota y_{i,0})), \\
\frac{\partial^2 l(\theta_0, \sigma_i^2)}{\partial \breve{\pi}^2} &= -\sum_{i=1}^{N} (\frac{1}{\sigma_i^2} (\iota y_{i,0})' \Phi^{-1} (\iota y_{i,0})), \\
\frac{\partial^2 l(\theta_0, \sigma_i^2)}{\partial \rho \partial \xi} &= \sum_{i=1}^{N} \left(\frac{1}{\sigma_i^2} u_i' \frac{\partial \Phi^{-1}}{\partial \xi} (\widetilde{\Delta} y_{i,-1}) \right) \\
&= -\sum_{i=1}^{N} \left(\frac{1}{\sigma_i^2} (\frac{\partial vec \Phi}{\partial \xi'})' \Phi^{-1} \otimes \Phi^{-1} vec(\widetilde{\Delta} y_{i,-1} u_i') \right), \\
\frac{\partial^2 l(\theta_0, \sigma_i^2)}{\partial \rho \partial \breve{\pi}} &= \sum_{i=1}^{N} \left(\frac{1}{\sigma_i^2} u_i' \frac{\partial \Phi^{-1}}{\partial \xi} (\iota y_{i,0}) \right), \\
\frac{\partial^2 l(\theta_0, \sigma_i^2)}{\partial \xi \partial \xi'} &= -\frac{1}{2} N (\frac{\partial vec \Phi}{\partial \xi'})' \Phi^{-1} \otimes \Phi^{-1} (\frac{\partial vec \Phi}{\partial \xi'}).
\end{aligned}
$$

$$
\begin{aligned}
plim_{N \to \infty} \frac{1}{N} \frac{\partial l(\theta_0)}{\partial \rho} \frac{\partial l(\theta_0)}{\partial \rho} &= plim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (\frac{1}{\sigma_i^4} (\widetilde{\Delta} y_{i,-1})' \Phi^{-1} u_i u_i' \Phi^{-1} (\widetilde{\Delta} y_{i,-1})), \\
plim_{N \to \infty} \frac{1}{N} \frac{\partial l(\theta_0)}{\partial \rho} \frac{\partial l(\theta_0)}{\partial \breve{\pi}} &= plim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (\frac{1}{\sigma_i^4} (\widetilde{\Delta} y_{i,-1})' \Phi^{-1} u_i u_i' \Phi^{-1} (\iota y_{i,0})),
\end{aligned}
$$

$$
\begin{aligned}
plim_{N\to\infty}\frac{1}{N}\frac{\partial l(\theta_0)}{\partial\breve{\pi}}\frac{\partial l(\theta_0)}{\partial\breve{\pi}} &= plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{\sigma_i^4}(\iota y_{i,0})'\Phi^{-1}u_iu_i'\Phi^{-1}(\iota y_{i,0})),\\
plim_{N\to\infty}\frac{1}{N}\frac{\partial l(\theta_0)}{\partial\rho}\frac{\partial l(\theta_0)}{\partial\xi} &= plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(-\frac{1}{2\sigma_i^4}(u_i'\Phi^{-1}\widetilde{\Delta}y_{i,-1})(u_i'\frac{\partial\Phi^{-1}}{\partial\xi}u_i)\right),\\
plim_{N\to\infty}\frac{1}{N}\frac{\partial l(\theta_0)}{\partial\breve{\pi}}\frac{\partial l(\theta_0)}{\partial\xi} &= plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(-\frac{1}{2\sigma_i^4}(u_i'\Phi^{-1}\iota y_{i,0})(u_i'\frac{\partial\Phi^{-1}}{\partial\xi}u_i)\right),\\
plim_{N\to\infty}\frac{1}{N}\frac{\partial l(\theta_0)}{\partial\xi}\frac{\partial l(\theta_0)}{\partial\xi'} &= plim_{N\to\infty}\frac{1}{2}(\frac{\partial vec\Phi}{\partial\xi'})'\Phi^{-1}\otimes\Phi^{-1}(\frac{\partial vec\Phi}{\partial\xi'}).
\end{aligned}
$$

Suppose the σ_i^2 are unknown: substitute $\widetilde{\sigma}_i^2=\widetilde{\sigma}_i^2(\theta_0)$ for σ_i^2 in $\frac{\partial l(\theta_0)}{\partial\theta_0}\frac{\partial l(\theta_0)}{\partial\theta_0'}$,

$$
\begin{aligned}
\frac{\partial\widetilde{\sigma}_i^2}{\partial\rho} &= -\frac{2}{T}(\widetilde{\Delta}y_{i,-1})'\Phi^{-1}u_i,\\
\frac{\partial\widetilde{\sigma}_i^2}{\partial\breve{\pi}} &= -\frac{2}{T}(\iota y_{i,0})'\Phi^{-1}u_i,\\
\frac{\partial\widetilde{\sigma}_i^2}{\partial\xi} &= -\frac{1}{T}(\frac{\partial vec\Phi}{\partial\xi'})'\Phi^{-1}\otimes\Phi^{-1}vec(u_iu_i').
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 l(\theta_0)}{\partial\rho^2} &= \frac{\partial^2 l(\theta_0,\widetilde{\sigma}_i^2)}{\partial\rho^2}-\sum_{i=1}^{N}(\frac{1}{\widetilde{\sigma}_i^4}(\widetilde{\Delta}y_{i,-1})'\Phi^{-1}u_i\frac{\partial\widetilde{\sigma}_i^2}{\partial\rho}),\\
\frac{\partial^2 l(\theta_0)}{\partial\rho\partial\breve{\pi}} &= \frac{\partial^2 l(\theta_0,\widetilde{\sigma}_i^2)}{\partial\rho\partial\breve{\pi}}-\sum_{i=1}^{N}(\frac{1}{\widetilde{\sigma}_i^4}(\iota y_{i,0})'\Phi^{-1}u_i\frac{\partial\widetilde{\sigma}_i^2}{\partial\rho}),\\
\frac{\partial^2 l(\theta_0)}{\partial\breve{\pi}^2} &= \frac{\partial^2 l(\theta_0,\widetilde{\sigma}_i^2)}{\partial\breve{\pi}^2}-\sum_{i=1}^{N}(\frac{1}{\widetilde{\sigma}_i^4}(\iota y_{i,0})'\Phi^{-1}u_i\frac{\partial\widetilde{\sigma}_i^2}{\partial\breve{\pi}}),\\
\frac{\partial^2 l(\theta_0)}{\partial\rho\partial\xi} &= \frac{\partial^2 l(\theta_0,\widetilde{\sigma}_i^2)}{\partial\rho\partial\xi}-\sum_{i=1}^{N}\left(\frac{1}{2\widetilde{\sigma}_i^4}(\frac{\partial vec\Phi^{-1}}{\partial\xi'})'vec(-u_iu_i')\frac{\partial\widetilde{\sigma}_i^2}{\partial\rho}\right),\\
\frac{\partial^2 l(\theta_0)}{\partial\breve{\pi}\partial\xi} &= \frac{\partial^2 l(\theta_0,\widetilde{\sigma}_i^2)}{\partial\breve{\pi}\partial\xi}-\sum_{i=1}^{N}\left(\frac{1}{2\widetilde{\sigma}_i^4}(\frac{\partial vec\Phi^{-1}}{\partial\xi'})'vec(-u_iu_i')\frac{\partial\widetilde{\sigma}_i^2}{\partial\breve{\pi}}\right),\\
\frac{\partial^2 l(\theta_0)}{\partial\xi\partial\xi'} &= \frac{\partial^2 l(\theta_0,\widetilde{\sigma}_i^2)}{\partial\xi\partial\xi'}-\sum_{i=1}^{N}\left(\frac{1}{2\widetilde{\sigma}_i^4}(\frac{\partial vec\Phi}{\partial\xi'})'\Phi^{-1}\otimes\Phi^{-1}vec(u_iu_i')\frac{\partial\widetilde{\sigma}_i^2}{\partial\xi}\right).
\end{aligned}
$$

References

[1] Ahn, S.C., and P. Schmidt, 1997, Efficient estimation of dynamic panel data models: alternative assumptions and simplified estimation, Journal of econometrics 76, 309-321.

[2] Alvarez, J. and M. Arellano, 2003, The time series and cross-section asymptotics of dynamic panel data estimators. Econometrica, 71, 1121-1159.

[3] Alvarez, J., and M. Arellano, 2004, Robust likelihood estimation of dynamic panel data models, unpublished manuscript, University of Alicante and CEMFI, Madrid.

[4] Anderson, T.W., and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, Journal of econometrics 18, 47-82.

[5] Arellano, M., and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of economic studies 58, 277-297.

[6] Bai, J., 2013, Fixed-effects dynamic panel models, a factor analytical method, Econometrica 81, 285-314.

[7] Bauwens, L., and J.V.K. Rombouts, 2007, Bayesian clustering of many GARCH models, Econometric Reviews 26, 365-286.

[8] Blundell, R.W., and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, Journal of econometrics 87, 115-143.

[9] Chamberlain, G., 1980, Analysis of covariance with qualitative data, Review of economic studies XLVII, 225-238.

[10] Chamberlain, G., 1982, Multivariate regression models for panel data, Journal of econometrics 18, 5-46.

[11] Chamberlain, G., 1984, Panel data, in Handbook of econometrics, edited by Z. Griliches and M. Intrilligator, North Holland, Amsterdam.

[12] Cragg, J.G., 1983, More efficient estimation in the presence of heteroskedasticity of unknown form, Econometrica 51, 751-763.

[13] Engle, R.F., and J. Mezrich, 1996, GARCH for groups, Risk 9, 36-40.

[14] Hansen, L.P, 1982, Large sample properties of GMM estimators, Econometrica 50, 1029-1054.

[15] Hayakawa, K. and M.H. Pesaran, 2012, Robust standard errors in transformed likelihood estimation of dynamic panel data models, unpublished manuscript, Hiroshima University, University of Cambridge and USC.

[16] Hsiao, C., M.H. Pesaran, and A.K. Tahmiscioglu, 2002, Maximum likelihood estimation of fixed effects dynamic panel models covering short time periods, Journal of econometrics 109, 107-150.

[17] Johnson, and S. Kotz, 1972, Distributions in statistics: continuous multivariate distributions, Wiley, New York.

[18] Kiefer, J., and J. Wolfowitz, 1956, Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, Annals of Mathematical Statistics 27, 887-906.

[19] Kiefer, N.M., 1980, Estimation of fixed effect models for time series of cross-sections with arbitrary intertemporal covariance, Journal of econometrics 14, 195-202.

[20] Kruiniger, H., 2008, Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model, Journal of econometrics 144, 447-464.

[21] Kruiniger, H., 2013, Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions, Journal of Econometrics 173, 175-188.

[22] Li, Q., and T. Stengos, 1994, Adaptive estimation in the panel data error component model with heteroskedasticity of unknown form, International economic review 35, 981-1000.

[23] MaCurdy, T., 1982, The use of time series processes to model the time structure of earnings in a longitudinal data analysis, Journal of econometrics 18, 83-114.

[24] Magnus, J.R., 1982, Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood, Journal of Econometrics 19, 239-285.

[25] Magnus, J.R., and H. Neudecker, 1988, Matrix differential calculus with applications in statistics and econometrics, Wiley, Chicester, United Kingdom.

[26] Meghir, C., and F. Windmeijer, 1999, Moment conditions for dynamic panel data models with multiplicative individual effects in the conditional variance, Annales d'économie et de statistique, N° 55-56, 317-330.

[27] Newey, W., and D. McFadden, Large sample estimation and hypothesis testing, Ch. 36 in Handbook of econometrics, vol. 4, R.F. Engle and D. McFadden eds., Elsevier, Amsterdam.

[28] Neyman, J., and E.L. Scott, 1948, Consistent estimates based on partially consistent observations, Econometrica 16, 1-32.

[29] Pakel, C., N. Shephard and K. Sheppard, 2011, Nuisance Parameters, Composite Likelihoods and a Panel of GARCH Models, Statistica Sinica 21, 307-329.

[30] Phillips, P.C.B. and H. Moon, 1999, Linear Regression Limit Theory for Nonstationary Panel Data. Econometrica 67, 1057-1111.

[31] Robinson, P.M., 1987, Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form, Econometrica 55, 875-891.

[32] Stock, J.H. and M.W. Watson, 2008, Heteroskedasticity-robust standard errors for fixed effects panel data regression, Econometrica 76, 155-174.

[33] White, H., 1994, Estimation, inference and specification analysis, Econometric Society Monographs No. 22, Cambridge University Press.

Table 1: Estimators of ρ; Design I-S; 5000 replications.

$N = 100$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	-.0001	0.0837	.0016	0.0906	.0071	0.1241	.0060	0.1158
4	0.50	.0132	0.1288	.0324	0.1703	.0303	0.1860	.0427	0.2022
4	0.80	.0416	0.1766	.0388	0.1879	.0392	0.2119	.0402	0.2145
4	0.95	.0535	0.1825	.0388	0.1772	.0498	0.2163	.0477	0.2037
9	0.20	-.0002	0.0400	-.0003	0.0400	-.0006	0.0436	-.0006	0.0436
9	0.50	-.0007	0.0424	-.0007	0.0424	.0001	0.0480	-.0003	0.0480
9	0.80	.0072	0.0663	.0158	0.0854	.0095	0.0748	.0206	0.0964
9	0.95	.0182	0.0755	.0174	0.0812	.0183	0.0806	.0088	0.0872
24	0.20	.0001	0.0218	.0001	0.0218	-.0001	0.0228	-.0001	0.0228
24	0.50	-.0006	0.0206	-.0006	0.0206	-.0005	0.0214	-.0005	0.0214
24	0.80	-.0006	0.0184	-.0006	0.0188	-.0007	0.0193	-.0006	0.0197
24	0.95	-.0084	0.0167	.0006	0.0230	-.0089	0.0175	-.0015	0.0230

Table 2: Estimators of ρ; Design I-S; 5000 replications.

$N = 500$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0010	0.0361	.0012	0.0374	.0014	0.0480	.0015	0.0458
4	0.50	.0012	0.0469	.0028	0.0548	.0030	0.0671	.0049	0.0728
4	0.80	.0078	0.0794	.0270	0.1196	.0144	0.1063	.0298	0.1338
4	0.95	.0223	0.1091	.0097	0.1091	.0179	0.1334	.0126	0.1245
9	0.20	.0004	0.0174	.0004	0.0174	.0004	0.0193	.0004	0.0193
9	0.50	.0001	0.0190	.0001	0.0192	.0002	0.0209	.0002	0.0210
9	0.80	.0004	0.0238	.0015	0.0285	.0003	0.0268	.0019	0.0323
9	0.95	.0045	0.0397	.0101	0.0502	.0038	0.0428	.0077	0.0524
24	0.20	.	.	.	.	.	.	.	.
24	0.50	.	.	.	.	.	.	.	.
24	0.80	.	.	.	.	.	.	.	.
24	0.95	.	.	.	.	.	.	.	.

Table 3: Estimators of ρ; Design II-S; 5000 replications.

$N = 100$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0008	0.0894	.0020	0.0949	.0070	0.1237	.0065	0.1175
4	0.50	.0132	0.1360	.0280	0.1712	.0265	0.1803	.0374	0.1960
4	0.80	.0431	0.1844	.0436	0.1965	.0432	0.2128	.0374	0.2100
4	0.95	.0511	0.1871	.0387	0.1811	.0460	0.2163	.0432	0.2020
9	0.20	-.0009	0.0424	-.0008	0.0424	-.0004	0.0458	-.0003	0.0447
9	0.50	.0010	0.0458	.0008	0.0458	.0009	0.0480	.0005	0.0490
9	0.80	.0081	0.0714	.0169	0.0894	.0093	0.0742	.0189	0.0933
9	0.95	.0182	0.0775	.0197	0.0843	.0168	0.0806	.0116	0.0866
24	0.20	-.0001	0.0226	-.0001	0.0226	.0001	0.0225	.0001	0.0224
24	0.50	-.0006	0.0217	-.0006	0.0217	-.0007	0.0214	-.0007	0.0214
24	0.80	-.0010	0.0200	-.0010	0.0204	-.0009	0.0194	-.0009	0.0198
24	0.95	-.0090	0.0176	-.0002	0.0234	-.0086	0.0171	-.0010	0.0226

Table 4: Estimators of ρ; Design II-S; 5000 replications.

$N = 500$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0012	0.0376	.0013	0.0387	.0022	0.0481	.0016	0.0452
4	0.50	.0011	0.0500	.0034	0.0591	.0045	0.0672	.0057	0.0724
4	0.80	.0098	0.0850	.0279	0.1219	.0143	0.1066	.0301	0.1327
4	0.95	.0259	0.1145	.0127	0.1141	.0224	0.1339	.0138	0.1256
9	0.20	.0001	0.0188	.0001	0.0188	-.0002	0.0197	-.0002	0.0197
9	0.50	-.0007	0.0199	-.0007	0.0201	-.0006	0.0210	-.0007	0.0211
9	0.80	.0007	0.0254	.0024	0.0317	.0009	0.0270	.0027	0.0333
9	0.95	.0060	0.0423	.0111	0.0520	.0045	0.0440	.0064	0.0524
24	0.20	.	.	.	.	.	.	.	.
24	0.50	.	.	.	.	.	.	.	.
24	0.80	.	.	.	.	.	.	.	.
24	0.95	.	.	.	.	.	.	.	.

Table 5: Estimators of ρ; Design III-S; 5000 replications.

$N = 100$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0138	0.1634	.0199	0.1828	.0172	0.1772	.0227	0.1825
4	0.50	.0360	0.2189	.0521	0.2508	.0393	0.2326	.0548	0.2567
4	0.80	.0518	0.2425	.0662	0.2604	.0455	0.2494	.0471	0.2585
4	0.95	.0520	0.2364	.0520	0.2396	.0499	0.2466	.0511	0.2387
9	0.20	-.0015	0.0678	-.0016	0.0678	.0000	0.0624	.0000	0.0624
9	0.50	-.0008	0.0775	-.0002	0.0806	.0003	0.0700	.0011	0.0728
9	0.80	.0183	0.1068	.0350	0.1311	.0091	0.0933	.0242	0.1187
9	0.95	.0169	0.1015	.0205	0.1082	.0121	0.0959	.0039	0.1034
24	0.20	.0013	0.0364	.0013	0.0364	.0007	0.0329	.0007	0.0328
24	0.50	-.0017	0.0351	-.0018	0.0351	-.0018	0.0302	-.0018	0.0302
24	0.80	-.0024	0.0308	-.0008	0.0334	-.0022	0.0263	-.0012	0.0283
24	0.95	-.0140	0.0261	-.0028	0.0307	-.0126	0.0224	-.0042	0.0268

Table 6: Estimators of ρ; Design III-S; 5000 replications.

$N = 500$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0013	0.0608	.0020	0.0632	.0019	0.0693	.0021	0.0663
4	0.50	.0067	0.0917	.0166	0.1233	.0102	0.1020	.0187	0.1245
4	0.80	.0288	0.1400	.0462	0.1712	.0180	0.1378	.0291	0.1612
4	0.95	.0438	0.1584	.0312	0.1543	.0325	0.1631	.0249	0.1565
9	0.20	-.0009	0.0310	-.0009	0.0311	-.0001	0.0293	-.0000	0.0292
9	0.50	-.0003	0.0326	-.0003	0.0329	.0001	0.0310	.0002	0.0312
9	0.80	.0016	0.0454	.0100	0.0647	.0009	0.0408	.0083	0.0610
9	0.95	.0118	0.0612	.0178	0.0698	.0063	0.0562	.0091	0.0665
24	0.20	.	.	.	.	.	.	.	.
24	0.50	.	.	.	.	.	.	.	.
24	0.80	.	.	.	.	.	.	.	.
24	0.95	.	.	.	.	.	.	.	.

Table 7: Estimators of ρ; Design I-NS; 5000 replications.

$N = 100$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0048	0.1034	.0050	0.1039	.0104	0.1342	.0102	0.1319
4	0.50	.0265	0.1609	.0261	0.1600	.0401	0.2040	.0363	0.1967
4	0.80	.0572	0.2059	.0569	0.2035	.0522	0.2330	.0268	0.2152
4	0.95	.0442	0.1841	.0354	0.1780	.0516	0.2182	.0291	0.1985
9	0.20	-.0003	0.0436	-.0004	0.0436	-.0002	0.0480	-.0002	0.0480
9	0.50	-.0011	0.0469	-.0012	0.0469	-.0004	0.0520	-.0005	0.0520
9	0.80	.0113	0.0812	.0106	0.0806	.0136	0.0889	.0123	0.0877
9	0.95	.0201	0.0819	.0164	0.0806	.0098	0.0860	-.0006	0.0843
24	0.20	-.0053	0.0221	-.0051	0.0220	-.0055	0.0230	-.0052	0.0230
24	0.50	-.0053	0.0197	-.0050	0.0196	-.0055	0.0205	-.0052	0.0204
24	0.80	-.0069	0.0170	-.0065	0.0169	-.0072	0.0175	-.0068	0.0175
24	0.95	-.0125	0.0209	-.0122	0.0209	-.0129	0.0215	-.0125	0.0215

Table 8: Estimators of ρ; Design I-NS; 5000 replications.

$N = 500$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0014	0.0424	.0014	0.0424	.0010	0.0520	.0009	0.0520
4	0.50	.0044	0.0640	.0044	0.0640	.0082	0.0831	.0078	0.0825
4	0.80	.0341	0.1296	.0319	0.1281	.0376	0.1473	.0155	0.1349
4	0.95	.0335	0.1166	.0290	0.1149	.0268	0.1342	.0044	0.1245
9	0.20	.0001	0.0190	.0001	0.0190	.0003	0.0207	.0003	0.0207
9	0.50	-.0005	0.0207	-.0005	0.0206	-.0005	0.0227	-.0005	0.0227
9	0.80	.0013	0.0311	.0013	0.0311	.0022	0.0356	.0021	0.0353
9	0.95	.0143	0.0525	.0128	0.0520	.0109	0.0547	.0028	0.0532
24	0.20	.	.	.	.	.	.	.	.
24	0.50	.	.	.	.	.	.	.	.
24	0.80	.	.	.	.	.	.	.	.
24	0.95	.	.	.	.	.	.	.	.

Table 9: Estimators of ρ; Design III-NS; 5000 replications.

$N = 100$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0155	0.1892	.0147	0.1881	.0174	0.1934	.0174	0.1895
4	0.50	.0329	0.2468	.0335	0.2456	.0343	0.2504	.0342	0.2449
4	0.80	.0408	0.2585	.0384	0.2550	.0351	0.2676	.0122	0.2522
4	0.95	.0465	0.2423	.0354	0.2371	.0615	0.2561	.0465	0.2377
9	0.20	-.0026	0.0714	-.0026	0.0714	-.0011	0.0663	-.0007	0.0656
9	0.50	.0015	0.0812	.0013	0.0812	-.0009	0.0755	-.0001	0.0762
9	0.80	.0161	0.1187	.0168	0.1192	.0118	0.1105	.0135	0.1105
9	0.95	.0127	0.1077	.0082	0.1072	.0060	0.1039	-.0032	0.1034
24	0.20	-.0061	0.0360	-.0055	0.0359	-.0052	0.0312	-.0047	0.0312
24	0.50	-.0094	0.0333	-.0086	0.0333	-.0082	0.0294	-.0076	0.0294
24	0.80	-.0133	0.0292	-.0123	0.0291	-.0107	0.0246	-.0100	0.0245
24	0.95	-.0200	0.0332	-.0193	0.0331	-.0183	0.0291	-.0174	0.0288

Table 10: Estimators of ρ; Design III-NS; 5000 replications.

$N = 500$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0015	0.0742	.0015	0.0742	.0011	0.0787	.0012	0.0787
4	0.50	.0142	0.1229	.0143	0.1233	.0155	0.1285	.0155	0.1273
4	0.80	.0483	0.1794	.0460	0.1780	.0382	0.1808	.0166	0.1676
4	0.95	.0331	0.1591	.0280	0.1559	.0358	0.1643	.0182	0.1520
9	0.20	.0006	0.0322	.0006	0.0322	.0003	0.0302	.0004	0.0302
9	0.50	.0002	0.0357	.0002	0.0357	-.0001	0.0335	.0000	0.0335
9	0.80	.0058	0.0610	.0057	0.0607	.0042	0.0547	.0048	0.0553
9	0.95	.0127	0.0698	.0114	0.0692	.0089	0.0673	.0009	0.0666
24	0.20	.	.	.	.	.	.	.	.
24	0.50	.	.	.	.	.	.	.	.
24	0.80	.	.	.	.	.	.	.	.
24	0.95	.	.	.	.	.	.	.	.

Table 11: Estimators of ρ; Design IV-S; 5000 replications.

$N = 100$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0121	.1327	.0172	.1523	.0000	.1386	-.0013	.1510
4	0.50	.0281	.1688	.0377	.1942	.0746	.2550	.0722	.2655
4	0.80	.0644	.1947	.0982	.2186	.2191	.3175	.2789	.3569
4	0.95	.0816	.1972	.0988	.2064	.1513	.2410	.1763	.2602
9	0.20	.0040	.0583	.0027	.0592	-.0003	.0640	-.0004	.0656
9	0.50	.0079	.0608	.0054	.0624	-.0013	.0678	-.0023	.0671
9	0.80	.0332	.0843	.0319	.0900	.0215	.0922	.0275	.1015
9	0.95	.0434	.0794	.0490	.0854	.0128	.0548	.0137	.0566
24	0.20	.0022	.0306	.0015	.0305	.0023	.0346	.0021	.0348
24	0.50	.0033	.0287	.0019	.0286	.0005	.0335	.0004	.0334
24	0.80	.0070	.0265	.0044	.0255	-.0009	.0277	-.0004	.0294
24	0.95	.0148	.0283	.0156	.0303	.0002	.0103	.0002	.0103

Table 12: Estimators of ρ; Design IV-S; 5000 replications.

$N = 500$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0016	.0574	.0014	.0600	-.0048	.0583	-.0051	.0632
4	0.50	.0133	.0927	.0122	.0964	.0028	.0854	.0021	.0877
4	0.80	.0499	.1432	.0879	.1778	.2590	.3473	.3130	.3837
4	0.95	.0756	.1581	.1036	.1764	.0894	.1803	.1047	.1942
9	0.20	.0010	.0284	.0005	.0285	-.0014	.0377	-.0016	.0391
9	0.50	.0013	.0302	.0004	.0303	-.0028	.0337	-.0029	.0345
9	0.80	.0136	.0524	.0063	.0370	.0070	.0543	.0074	.0558
9	0.95	.0319	.0648	.0467	.0801	.0004	.0149	.0004	.0149
24	0.20	.	.	.	.	.	.	.	.
24	0.50	.	.	.	.	.	.	.	.
24	0.80	.	.	.	.	.	.	.	.
24	0.95	.	.	.	.	.	.	.	.

Table 13: Estimators of ρ; Design IV-NS; 5000 replications.

$N = 100$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	0.0210	0.152	0.0208	0.1513	0.0042	0.1838	0.0051	0.1797
4	0.50	0.0421	0.200	0.0407	0.1975	0.0905	0.2890	0.0877	0.2877
4	0.80	0.0786	0.214	0.0801	0.2114	0.2375	0.3297	0.2790	0.3557
4	0.95	0.0945	0.208	0.0991	0.2049	0.1633	0.2502	0.1860	0.2653
9	0.20	0.0121	0.062	0.0124	0.0624	0.0045	0.0742	0.0046	0.0721
9	0.50	0.0164	0.069	0.0167	0.0686	-0.0045	0.0768	-0.0043	0.0748
9	0.80	0.0445	0.098	0.0450	0.0985	0.0318	0.1095	0.0343	0.1118
9	0.95	0.0498	0.085	0.0507	0.0849	0.0122	0.0548	0.0134	0.0566
24	0.20	0.0065	0.0324	0.0065	0.0324	0.0019	0.0351	0.0019	0.0348
24	0.50	0.0057	0.0301	0.0058	0.0302	-0.0007	0.0350	-0.0005	0.0347
24	0.80	0.0109	0.0298	0.0111	0.0299	-0.0003	0.0290	0.0001	0.0305
24	0.95	0.0191	0.0328	0.0196	0.0331	0.0002	0.0104	0.0002	0.0104

Table 14: Estimators of ρ; Design IV-NS; 5000 replications.

$N = 500$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0069	.0693	.0069	.0693	-.0056	.0735	-.0055	.0728
4	0.50	.0183	.1118	.0178	.1105	.0063	.1105	.0035	.1005
4	0.80	.0645	.1628	.0629	.1603	.2856	.3648	.3020	.3754
4	0.95	.0956	.1718	.0931	.1667	.0999	.1905	.1055	.1954
9	0.20	.0032	.0306	.0032	.0306	-.0008	.0406	-.0008	.0403
9	0.50	.0038	.0332	.0038	.0332	-.0024	.0371	-.0024	.0370
9	0.80	.0149	.0531	.0151	.0530	.0089	.0597	.0057	.0548
9	0.95	.0478	.0808	.0478	.0801	.0003	.0146	.0004	.0148
24	0.20	.	.	.	.	.	.	.	.
24	0.50	.	.	.	.	.	.	.	.
24	0.80	.	.	.	.	.	.	.	.
24	0.95	.	.	.	.	.	.	.	.

Table 15: Estimators of ρ; Design V-S; 5000 replications.

$N = 100$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0013	.0781	.0019	.0819	.0068	.1187	.0041	.1086
4	0.50	.0056	.1039	.0140	.1304	.0237	.1682	.0400	.1960
4	0.80	.0216	.1503	.0418	.1871	.0500	.2090	.1368	.2837
4	0.95	.0391	.1811	.0430	.1895	.0919	.2339	.1580	.2623
9	0.20	-.0009	.0400	-.0009	.0400	-.0009	.0436	-.0008	.0436
9	0.50	.0003	.0424	.0002	.0424	.0003	.0469	.0003	.0469
9	0.80	.0010	.0500	.0035	.0592	.0060	.0640	.0172	.0877
9	0.95	.0160	.0735	.0248	.0843	.0294	.0849	.0593	.1095
24	0.20	.0004	.0218	.0004	.0218	.0006	.0229	.0006	.0229
24	0.50	-.0001	.0206	-.0002	.0206	-.0001	.0212	-.0002	.0213
24	0.80	-.0019	.0180	-.0015	.0181	-.0020	.0189	-.0015	.0190
24	0.95	-.0053	.0141	-.0015	.0186	-.0053	.0148	-.0015	.0195

Table 16: Estimators of ρ; Design V-S; 5000 replications.

$N = 500$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	-.0002	.0346	.0000	.0346	-.0005	.0458	-.0000	.0424
4	0.50	.0002	.0412	.0007	.0436	.0017	.0566	.0017	.0539
4	0.80	.0031	.0574	.0098	.0806	.0098	.0900	.0526	.1655
4	0.95	.0157	.1000	.0241	.1166	.0520	.1597	.1118	.2066
9	0.20	-.0001	.0174	-.0002	.0174	-.0002	.0193	-.0002	.0192
9	0.50	.0001	.0185	.0001	.0186	.0001	.0206	.0002	.0204
9	0.80	-.0000	.0209	.0000	.0214	.0003	.0237	.0006	.0258
9	0.95	.0046	.0349	.0074	.0427	.0092	.0440	.0503	.0920
24	0.20	.	.	.	.	.	.	.	.
24	0.50	.	.	.	.	.	.	.	.
24	0.80	.	.	.	.	.	.	.	.
24	0.95	.	.	.	.	.	.	.	.

Table 17: Estimators of ρ; Design VI-S; 5000 replications.

$N = 100$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0071	.1490	.0100	.1606	.0164	.1783	.0226	.1857
4	0.50	.0214	.1926	.0366	.2241	.0360	.2232	.0715	.2740
4	0.80	.0375	.2245	.0583	.2506	.0597	.2470	.1526	.3156
4	0.95	.0503	.2335	.0610	.2421	.0909	.2604	.1475	.2818
9	0.20	-.0001	.0671	-.0001	.0671	-.0005	.0624	-.0003	.0624
9	0.50	.0001	.0735	.0002	.0742	-.0009	.0678	-.0006	.0686
9	0.80	.0098	.0943	.0236	.1170	.0067	.0849	.0268	.1192
9	0.95	.0247	.0970	.0342	.1077	.0258	.0964	.0513	.1175
24	0.20	-.0008	.0359	-.0008	.0359	-.0008	.0327	-.0008	.0327
24	0.50	-.0016	.0348	-.0016	.0348	-.0010	.0300	-.0010	.0300
24	0.80	-.0025	.0296	-.0015	.0311	-.0012	.0258	-.0008	.0267
24	0.95	-.0126	.0251	-.0035	.0291	-.0099	.0200	-.0024	.0248

Table 18: Estimators of ρ; Design VI-S; 5000 replications.

$N = 500$	$\sigma_\mu^2 = 1$	REUQMLE		FEUQMLE		REWQMLE		FEWQMLE	
T	ρ	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
4	0.20	.0010	.0592	.0013	.0608	.0002	.0648	.0007	.0624
4	0.50	.0027	.0755	.0054	.0866	.0055	.0860	.0123	.1105
4	0.80	.0144	.1145	.0332	.1543	.0177	.1249	.0945	.2263
4	0.95	.0337	.1562	.0370	.1667	.0715	.1918	.1379	.2369
9	0.20	.0000	.0303	.0000	.0303	.0004	.0289	.0004	.0287
9	0.50	.0002	.0323	.0002	.0324	.0004	.0294	.0004	.0295
9	0.80	.0009	.0377	.0020	.0416	.0013	.0367	.0059	.0528
9	0.95	.0130	.0608	.0210	.0735	.0142	.0599	.0578	.1034
24	0.20	.	.	.	.	.	.	.	.
24	0.50	.	.	.	.	.	.	.	.
24	0.80	.	.	.	.	.	.	.	.
24	0.95	.	.	.	.	.	.	.	.