Quasi ML estimation of the panel AR(1) model with additional regressors

Hugo Kruiniger*
Durham University

This version: 5 March 2017

JEL classification: C12, C13, C23.

Keywords: correlated effects, dynamic panel data, factor structure, FIML, fixed effects, Generalized Method of Moments (GMM), heteroskedasticity, initial conditions, interactive effects, LIML, predetermined regressors, Quasi Maximum Likelihood (QML), strict exogeneity, weak exogeneity.

*Address: hugo.kruiniger@durham.ac.uk; Department of Economics, 1 Mill Hill Lane, Durham DH1 3LB, England.

Abstract

In this paper we discuss several limited information (LI) and full information (FI) random effects and fixed effects Quasi ML estimators (MLEs) for panel AR(1) models with additional regressors. We also consider related GMM estimators. All estimators are consistent for short (large N, fixed T) panels. The models allow for arbitrary initial conditions and heteroskedasticity and are extensions and generalizations of the models considered in Kruiniger (2013. Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions. Journal of Econometrics 173, 175-188). Among other things, we distinguish between the case where the regressors are strictly exogenous, the case where some of them are predetermined with respect to the idiosyncratic errors, including the case where they are weakly exogenous, and the case where some regressors are contemporaneously correlated with the idiosyncratic errors; we consider the possibility that the regressors are correlated with the individual effects; and we discuss estimation of models with time-varying individual effects. We also discuss how to choose between a random effects and a fixed effects approach. When the distribution of the data is correctly specified, the LI MLEs have better finite sample properties than the corresponding GMM estimators and when the time-dimension, T, is not small relative to the cross-section dimension, N, Wald tests based on the QMLEs have better size properties than GMM based Wald tests. Finally, the LI QMLEs are more easily computed and are often more precise than the FI QMLEs.

1 Introduction

In this paper we discuss consistent Random Effects (RE) and Fixed Effects (FE) Quasi ML estimators and related GMM estimators for variations of the following panel AR(1) model with one additional regressor:[1] [2]

$$y_{i,t} = \rho y_{i,t-1} + \beta(1-\rho)x_{i,t} + (1-\rho)\mu_i + \varepsilon_{i,t}, \tag{1}$$

for $i = 1, ..., N$ and $t = 2, ..., T$. We assume that $x_{i,1}$, $i = 1, ..., N$ are also observed. The claims that we will make about the properties of the estimators are based on large N, fixed T asymptotics. The FE estimators only exploit data in first differences and hence can rely on *minimal* assumptions for their consistency whereas the RE estimators exploit data in levels and are often more efficient than their FE counterparts. The models that we consider allow for arbitrary initial conditions and heteroskedasticity and are extensions and generalizations of the basic panel AR(1) models outlined in e.g. Kruiniger (2013). Specifically, we distinguish between the case where the regressor $x_{i,t}$ is strictly exogenous with respect to the idiosyncratic errors $\varepsilon_{i,s}$, i.e., $E(\varepsilon_{i,t}|y_i^{t-1}, x_i^T, \mu_i) = 0$, $t = 2, ..., T$, where $y_i^t = (y_{i,1} \ ... \ y_{i,t})'$ and $x_i^t = (x_{i,1} \ ... \ x_{i,t})'$, the case where $x_{i,t}$ is predetermined with respect to $\varepsilon_{i,t}$, i.e., $E(\varepsilon_{i,t}|y_i^{t-1}, x_i^t, \mu_i) = 0$, $t = 2, ..., T$, including the case where $x_{i,t}$ is weakly exogenous with respect to the parameters ρ and β, and the case where some of the regressors are contemporaneously correlated with the idiosyncratic errors. In addition, we allow for the possibility that $x_{i,t}$ is correlated with the individual effect μ_i. We also consider models with time-varying individual effects, i.e., with a factor structure, cf. e.g. Holtz-Eakin et al. (1988) and Ahn et al. (2001, 2013). Finally we discuss both Limited Information (LI) and Full Information (FI) QMLEs. The QMLEs allow for different variance parameters over time. In most applications this is crucial to ensure their consistency.

In section 2 we first discuss the QML and GMM estimators for models with strictly exogenous regressors. Our FEMLE for the model with time-invariant individual effects is a generalization of an MLE of Hsiao et al. (2002) and our FEMLE for the model with time-varying individual effects is a generalization of an MLE of Hayakawa et al. (2014). We also prove consistency of the FEQMLEs without invoking untestable assumptions. Next we propose new (single equation based) LI QMLEs for various models with predeter-

[1] Extensions to models with multiple additional regressors are straightforward.

[2] A constant, additive time dummies and time trend can easily be included but have been omitted to keep the exposition simple.

mined regressors that depend on (some of) the same individual effect(s) as the dependent variable does. Moral-Benito (2013) discusses what he calls subsystem LIMLEs (ssLIMLEs) and Bai (2013b) discusses FI QMLEs for such models; Bai (2013b) only considers LI QMLEs for models with predetermined regressors that are weakly exogenous.[3] Furthermore, unlike Bai (2013b) and Moral-Benito, we also suggest FE estimators and, unlike Moral-Benito, we also propose estimators for models with time-varying individual effects.

Moral-Benito's ssLIMLE is based on a model that consists of $T-1$ structural equations for the dependent variable completed with a set of unrestricted reduced form equations for the initial observations and the k predetermined regressors for periods $2, ..., T$. The reduced form equations for the predetermined regressors, e.g. the $x_{i,t}$, are period-specific linear projections of the $x_{i,t}$ on all available lags, i.e., y_i^{t-1} and x_i^{t-1}, for $t = 2, ..., T$. [4] Moral-Benito also allows the first two moments of the joint distribution of the initial observations and the individual effects to be unrestricted. This parametrization of the model is labeled as the Full Covariance Structure representation and potentially leads to a very large number of parameters, namely $O(T^2k^2)$, and burdensome computation of the ssLIMLE. Moral-Benito also presents an alternative Simultaneous Equations representation of the model, which leads to a profile likelihood function that depends on $O(T^2k)$ parameters and whose maximum is more easily computed. Our LI (Q)MLEs for models with predetermined regressors are obtained by estimating $T-1$ augmented structural equations for the dependent variable which include residuals from structural equations for the predetermined regressors, that is, by estimating $k+1$ models that each contain $T-1$ equations and $O(T)+O(k)$ parameters. As a result our LI(Q)MLEs are more easily computed than the ssLIMLE.[5]

In section 2 we also discuss FI estimators, like Bai (2013b). Bai (2013b) also develops inferential theory for panels with large T. The large T analysis is challenging and nonstandard due to the incidental variance parameters in the time-dimension, cf. Bai (2013a).

[3]The MLE of Hsiao et al. (2002) for such models appears to be inconsistent, see also remark 3.

[4]Practically speaking, in this feedback process the highest *common* lag length *across time* is only one for both y and x just as in (9); the real potential benefit of using these projections is that they ensure that the ssLIMLE is consistent even if the true model for the predetermined regressors contains additional (time-varying) individual effects, cf. Arellano (2016). However, in the latter case Moral-Benito's model can no longer be used to test for strict exogeneity.

[5]The approach of Moral-Benito can also be simplified by replacing the reduced form equations for the predetermined variables by structural equations. This would result in a FIQMLE. However, our LIQMLEs would still be easier to compute, especially when the number of predetermined regressors is large.

As our focus is on estimation procedures rather than the challenges of large T inference, we only consider large N, fixed T consistency of the estimators for the sake of brevity.

In section 3 we discuss tests of RE versus FE specifications.

If the distribution of the data is correctly specified, then the MLEs will have better finite sample properties than their GMM counterparts, also when the instruments are weak and/or many, cf. Anderson et al. (1982), Alvarez and Arellano (2003), Kruiniger (2013) and Hsiao and Zhang (2015). However, when the distribution of the data is incorrectly specified, then the Quasi ML estimators will generally asymptotically be less efficient than the optimal GMM estimators. In section 4 we examine the finite sample properties of some of the estimators and related Wald tests in a Monte Carlo study.

2 The models and the estimators

Throughout the paper we rely on the following **Basic Assumptions**: (i) $-1 < \rho \leq 1$. (ii) $T \geq 3$ (iii) The observations are independently distributed across the individuals conditional on the initial observations and, if present, factor(s). (iv) $E(\varepsilon_{i,t}|y_i^{t-1}, x_i^{t-1}, \mu_i) = 0$ for $i = 1, ..., N$ and $t = 2, ..., T$. (v) Relevant moments of the data exist as required for establishing the asymptotic properties of the estimators. E.g., consistency of the QML (optimal GMM) estimators requires existence of $(2+\epsilon)^{th}$ ($(4+\epsilon)^{th}$) moments, with $\epsilon > 0$.

Note that the regressor $x_{i,t}$ and the individual effect μ_i drop out from (1) when $\rho = 1$. The parametrization $(1-\rho)\mu_i$ in (1) prevents μ_i from turning into an individual trend at $\rho = 1$ and thereby avoids a discontinuity in the data generating process at $\rho = 1$. A similar comment applies to $\beta(1-\rho)x_{i,t}$ in (1). Although the MLEs that we discuss below are based on Gaussian likelihood functions, the true distributions of the data can be non-Gaussian and heterogeneous. In particular, the idiosyncratic errors are allowed to exhibit arbitrary heteroskedasticity across both dimensions of the panel even though the estimators that we propose use the same variance parameters for all individuals, cf. Kruiniger (2013). The assumptions also allow the errors to be conditionally heteroskedastic over time.

In this section we will first discuss QMLEs that are based on a single augmented equation, e.g. LI QMLEs, and subsequently we will discuss estimators that are based on a complete system of augmented equations, i.e., FI QMLEs. Related GMM estimators will be based on the score vectors. The consistency proofs for the LI QMLEs are straightforward generalizations of those given in the working paper (wp) version of Kruiniger (2013) for QMLEs for similar panel AR(1) models with time-invariant individual effects but without covariates. These proofs are reproduced in Appendix A below. [6] When $\rho = 1$, consistency of the QMLEs requires that $T \geq 4$ rather than $T \geq 3$, cf. Kruiniger (2013). Consistency of the FI QMLEs can be shown similarly, cf. Bai (2013b).

2.1 Limited information QML estimators

2.1.1 Dynamic panel models with strictly exogenous regressors

We first consider the case where the $x_{i,t}$ are strictly exogenous with respect to the $\varepsilon_{i,s}$, i.e., $E(\varepsilon_{i,t}|y_i^{t-1}, x_i^T, \mu_i) = 0$, $t = 2, ..., T$. To obtain a consistent LI REQMLE for ρ and β in (1), one can follow the approach of Mundlak (1978) and Chamberlain (1982) to remove the correlation between the regressors and the unobserved effects and estimate an augmented version of (1) in which μ_i is replaced by its linear projection on 1, $y_{i,1}$ and possibly other terms, and a new individual effect, v_i: [7]

$$\mu + \pi y_{i,1} \; \{+\varphi x_{i,1} + \gamma' x_i\} + v_i, \tag{2}$$

where $x_i = (x_{i,2} \; ... \; x_{i,T})'$, cf. Anderson and Hsiao (1982) and Bai (2013b). If $Cov(x_{i,t}, \mu_i - \pi y_{i,1}) = \sigma_{vx} \neq 0$, $t = 1, ..., T$, for some σ_{vx}, then consistency of the REQMLE requires adding a term like $\varphi x_{i,t}$, e.g. $\varphi x_{i,1}$, to the model. If $Cov(x_{i,t}, \mu_i - \pi y_{i,1}) \neq 0$ and $Cov(x_{i,s}, \mu_i - \pi y_{i,1}) \neq Cov(x_{i,t}, \mu_i - \pi y_{i,1})$ for (some) $s, t \in \{1, ..., T\}$ with $s \neq t$, then also adding the term $(1 - \rho)\gamma' x_i$ to the model again ensures consistency of the REQMLE.

Let $u_i = (1 - \rho) v_i \iota + \varepsilon_i$, where $\iota = (1 \; 1 \; ... \; 1)'$ and $\varepsilon_i = (\varepsilon_{i,2} \; ... \; \varepsilon_{i,T})'$. An optimal GMM estimator that exploits *at least* the moment conditions that are based on the score vector of the REMLE, viz. $E(y_{i,-1}' \Phi^{-1} u_i) = 0$, $E(x_i' \Phi^{-1} u_i) = 0$, $E(\iota' \Phi^{-1} u_i) = 0$, $E(y_{i,1} \iota' \Phi^{-1} u_i) = 0$, $\{E(x_{i,1} \iota' \Phi^{-1} u_i) = 0, E(x_i \iota' \Phi^{-1} u_i) = 0\}$, $E(tr((u_i u_i' - \Phi) \partial \Phi^{-1} / \partial \sigma_t^2)) = 0$, $t = 2, ..., T$, and $E(tr((u_i u_i' - \Phi) \partial \Phi^{-1} / \partial \widetilde{\sigma}_v^2)) = 0$, where $y_{i,-1} = (y_{i,1} \; ... \; y_{i,T-1})'$ and

[6] Consistency of the QMLEs can be proved by using Theorem 2.1 in Newey and McFadden (1994, NMcF). The QMLEs are only functions of the first two moments of the data. Assuming existence of $(2 + \epsilon)^{th}$ moments of the data if the data are heterogeneously distributed, where $\epsilon > 0$, one can show that the quasi likelihood functions converge uniformly in probability to the same non-random functions as they'd converge to if the data were i.i.d. and normal. Therefore to verify the other conditions of Theorem 2.1 of NMcF we can use Theorem 2.5 in NMcF. As we allow for heteroskedasticity of the errors, we can prove that the parameters are identified along the lines of Kruiniger (2013wp), see Appendix A below. The other conditions of Theorem 2.5 of NMcF, including the dominance condition, are easily verified, again see Appendix A.

[7] Below (some of) the terms or moment conditions in curly brackets are optional and their inclusion depends on the circumstances.

$\Phi = \widetilde{\sigma}_v^2 \iota\iota' + \Psi$ with $\widetilde{\sigma}_v^2 = (1-\rho)^2 \sigma_v^2$, $\sigma_v^2 = E(v_i^2)$ and $\Psi = diag(\sigma_2^2, ..., \sigma_T^2)$, is asymptotically at least as precise as the REQMLE, cf. Kruiniger (2013). Both estimators are asymptotically equivalent in special cases such as exact identification or when the data are i.i.d. across the individuals and Gaussian. Additional moment conditions that can be used to construct potentially more efficient GMM estimators are Arellano and Bond (1991) and Ahn and Schmidt (1995) type moment conditions.

To obtain a FEQMLE for ρ and β in (1), one can estimate a modified version of (1) in which μ_i is replaced by

$$\underline{\mu} + y_{i,1} - \beta x_{i,1} + \underline{v}_i.$$

where $\underline{v}_i$ is a new individual effect. However, this FEQMLE will be consistent only if $Cov(\Delta x_{i,t}, \underline{v}_i) = 0$, $t = 2, ..., T$. If $Cov(\Delta x_{i,t}, \underline{v}_i) \neq 0$ for some or all $t \in \{2, ..., T\}$, as often cannot be ruled out, then a consistent FEQMLE can be obtained by replacing $\underline{v}_i$ in turn by its linear projection on 1 and the elements of $(x_i - x_{i,-1})$ and a new individual effect v_i and estimating the augmented model that results, cf. Hsiao et al. (2002):

$$\widetilde{\Delta} y_i = \rho \widetilde{\Delta} y_{i,-1} + \widetilde{\beta} \widetilde{\Delta} x_i + \widetilde{\mu} \iota + \widetilde{\gamma}'(x_i - x_{i,-1})\iota + u_i, \tag{3}$$

where $y_i = (y_{i,2} \ ... \ y_{i,T})'$, $\widetilde{\Delta} y_i = y_i - y_{i,1}\iota$, $\widetilde{\Delta} y_{i,-1} = y_{i,-1} - y_{i,1}\iota$, $\widetilde{\Delta} x_i = x_i - x_{i,1}\iota$, $x_{i-1} = (x_{i,1} \ ... \ x_{i,T-1})'$, $\widetilde{\beta} = \beta(1-\rho)$, $\widetilde{\mu} = (1-\rho)\mu$, $\widetilde{\gamma} = (1-\rho)\gamma$ and $u_i = (1-\rho)v_i + \varepsilon_i$. We can rewrite (3) as

$$R \widetilde{\Delta} y_i = \widetilde{\beta} \widetilde{\Delta} x_i + \widetilde{\mu} \iota + \widetilde{\gamma}'(x_i - x_{i,-1})\iota + u_i, \tag{4}$$

where $R = R(1, \rho)$ is a constant bi-diagonal matrix such that $R \widetilde{\Delta} y_i = \widetilde{\Delta} y_i - \rho \widetilde{\Delta} y_{i,-1}$. Since $\det(R) = 1$, the likelihood of u_i is the same as the likelihood of $\widetilde{\Delta} y_i$ (or y_i) given ($y_{i,1}$ and) $x_i - x_{i,-1}$. Consistency of the FEQMLE based on (3) now follows from Theorems 2.1 and 2.5 in NMcF, see footnote 6. The identification and dominance conditions of NMcF's Theorem 2.5 can be verified similarly to Kruiniger (2013wp). [8]

Instead of (1) we can consider a more general model with time-varying effects, viz.

$$y_{i,t} = \rho y_{i,t-1} + \beta(1-\rho)x_{i,t} + \delta_t(1-\rho)\mu_i + \varepsilon_{i,t}. \tag{5}$$

[8] Our analysis is different from that in Hsiao et al. (2002). In particular, we don't invoke their assumption (3.ii)$'$, which states that for some $m \geq 0$, $E(\Delta y_{i,-m+2} | \Delta x_{i,2}, \Delta x_{i,3}, ..., \Delta x_{i,T})$ is equal to the same constant across all individuals. Assumption (3.ii)$'$ is quite strong and restrictive and cannot be tested as it is an assumption about presample data.

The δ_t will be treated as parameters. The preceding discussions of RE and FE QML estimation of ρ and β remain relevant for this more general model. However, the RE-QMLE for (5) is now based on

$$y_{i,t} = \rho y_{i,t-1} + \widetilde{\beta} x_{i,t} + \widetilde{\delta}_t(\mu + \pi y_{i,1}\ \{+\varphi x_{i,1} + \gamma' x_i\}) + u_{i,t}, \tag{6}$$

where $\widetilde{\delta}_t = \delta_t(1-\rho)$ and $u_{i,t} = \widetilde{\delta}_t v_i + \varepsilon_{i,t}$ (cf. Bai, 2013b), while the FEQMLE for (5), which only depends on first differences of the data, is based on

$$\widetilde{\Delta} y_{i,t} = \rho \widetilde{\Delta} y_{i,t-1} + \widetilde{\beta}\widetilde{\Delta} x_{i,t} + \widetilde{\mu}_1 + \widetilde{\gamma}_1'(x_i - x_{i,-1}) + \widetilde{\delta}_t \mu_2\ \{+\widetilde{\delta}_t \gamma_2'(x_i - x_{i,-1})\} + u_{i,t}, \tag{7}$$

where $\widetilde{\Delta} y_{i,t} = y_{i,t} - y_{i,1}$, $\widetilde{\Delta} y_{i,t-1} = y_{i,t-1} - y_{i,1}$, $\widetilde{\Delta} x_{i,t} = x_{i,t} - x_{i,1}$, $\widetilde{\mu}_1 = (1-\rho)\underline{\mu}_1$, $\widetilde{\gamma}_1 = (1-\rho)\gamma_1$ and $u_{i,t} = (1-\rho)v_{1,i} + \widetilde{\delta}_t v_i + \varepsilon_{i,t}$. To derive (7) we first rewrote (5) as

$$\widetilde{\Delta} y_{i,t} = \rho \widetilde{\Delta} y_{i,t-1} + \widetilde{\beta}\widetilde{\Delta} x_{i,t} - (1-\rho)(y_{i,1} - \beta x_{i,1}) + \delta_t(1-\rho)\mu_i + \varepsilon_{i,t}. \tag{8}$$

Next we replaced $(1-\rho)(y_{i,1}-\beta x_{i,1})$ and μ_i by projections on 1 and $x_{i,t}-x_{i,t-1}$, $t = 2, ..., T$. To achieve identification we can impose $\sigma_v^2 = 1$ in both the RE and FE case. The likelihood of u_i in (7) is the same as the likelihood of $\widetilde{\Delta} y_i$ (or y_i) given ($y_{i,1}$ and) $x_i - x_{i,-1}$ and consistency of the FEQMLE based on (7) follows.[9] Note that the $\widetilde{\delta}_t$ parameters appear in both the mean equation and the covariance matrix of the composite errors. The QMLEs for this model can be computed by using the ECM algorithm discussed in Bai (2013b).

Some of the moment conditions that are exploited by the GMM estimators related to the QMLEs for (5) are different from those for (1). For instance, in the case of the RE GMM estimator, replace $E(tr((u_i u_i' - \Phi)\partial\Phi^{-1}/\partial\widetilde{\sigma}_v^2)) = 0$, $E(\iota'\Phi^{-1}u_i) = 0$, $E(y_{i,1}\iota'\Phi^{-1}u_i) = 0$, $E(x_{i,1}\iota'\Phi^{-1}u_i) = 0$ and $E(x_i\iota'\Phi^{-1}u_i) = 0$ by $E((\mu + \pi y_{i,1}\ \{+\varphi x_{i,1} + \gamma' x_i\})e_{t-1}'\Phi^{-1}u_i) - \frac{1}{2}E(tr((u_i u_i' - \Phi)\partial\Phi^{-1}/\partial\widetilde{\delta}_t)) = 0$, $t = 2, ..., T$, $E(\widetilde{\delta}'\Phi^{-1}u_i) = 0$, $E(y_{i,1}\widetilde{\delta}'\Phi^{-1}u_i) = 0$, $E(x_{i,1}\widetilde{\delta}'\Phi^{-1}u_i) = 0$ and $E(x_i\widetilde{\delta}'\Phi^{-1}u_i) = 0$, where e_t is the $t{-}th$ column of the identity matrix I_{T-1}, $\widetilde{\delta} = (\widetilde{\delta}_2, \widetilde{\delta}_3, \widetilde{\delta}_4, ..., \widetilde{\delta}_T)'$ and $\Phi = \widetilde{\delta}\widetilde{\delta}' + \Psi$. Ahn, Lee and Schmidt (2001) suggest additional moment conditions that can be used by GMM estimators for (5).

[9] Our analysis is different from that in Hayakawa et al. (2014). In particular, we don't invoke their assumption 5, which states that for some $m \geq 0$, $E(\Delta y_{i,-m+2}|\Delta x_{i,2}, \Delta x_{i,3}, ..., \Delta x_{i,T})$ is equal to the same constant across all individuals. Assumption 5 is quite strong and restrictive and cannot be tested as it is an assumption about presample data. Furthermore, we allow for correlation between the $\Delta x_{i,t}$ and μ_i, which is important when $x_{i,t}$ obeys e.g. (14) with $\widetilde{\beta}_x = 0$.

2.1.2 Dynamic panel models with predetermined regressors

Let us now consider single equation based LI QML estimators for ρ and β in (1) or (5) when the $x_{i,t}$ are merely predetermined with respect to the $\varepsilon_{i,t}$, i.e., $E(\varepsilon_{i,t}|y_i^{t-1}, x_i^t, \mu_i) = 0$, $t = 2, ..., T$, but the $x_{i,t}$ may still be directly affected by (some of) the same individual effect(s) as the $y_{i,t}$. Initially we will assume that the $x_{i,t}$ obey the following specification:

$$x_{i,t} = \alpha_x x_{i,t-1} + \beta_x y_{i,t-1} + \gamma_x \mu_i + \lambda_i + \xi_{i,t}, \tag{9}$$

where $|\alpha_x| \leq 1$, β_x and γ_x are parameters, μ_i and λ_i are independent and $\xi_{i,s}$ and $\varepsilon_{i,t}$ are independent for all s, t.[10] As $x_{i,t}$ depends on $y_{i,t-1}$, $x_{i,t}$ is correlated with lags of $\varepsilon_{i,t}$ and hence predetermined with respect to $\varepsilon_{i,t}$. Furthermore, $x_{i,t}$ is correlated with μ_i even if $\gamma_x = 0$. However, if $\gamma_x = 0$, then $x_{i,t}$ is weakly exogenous with respect to ρ and β in both (1) and (5). A consistent LI REQMLE for ρ and β when $\beta_x \neq 0$ but $\gamma_x = 0$ is given in Bai (2013b). If $\beta_x = 0$, then $x_{i,t}$ is strictly exogenous with respect to the $\varepsilon_{i,s}$.

We will now describe a single equation LI REQML approach to estimating ρ and β in (1) when both $\beta_x \neq 0$ and $\gamma_x \neq 0$ in (9) but the $\xi_{i,t}$ are homoskedastic over time. Following the logic of the FI REQML approach (cf. Bai, 2013b, and section 2.2 below) we first consider replacing $(1-\rho)\mu_i$ in (1) by its projection on 1, $y_{i,1}$ and $x_{i,1}$, i.e.,

$$(1-\rho)\mu_i = \widetilde{\mu}_y + \widetilde{\pi}_y y_{i,1} + \widetilde{\varphi}_y x_{i,1} + \widetilde{v}_{y,i},$$

where $\widetilde{v}_{y,i}$ is the projection residual. Applying the ML method to

$$y_{i,t} = \rho y_{i,t-1} + \widetilde{\beta} x_{i,t} + \widetilde{\mu}_y + \widetilde{\pi}_y y_{i,1} + \widetilde{\varphi}_y x_{i,1} + \widetilde{u}_{y,i,t},$$

with $\widetilde{u}_{y,i,t} = \widetilde{v}_{y,i} + \varepsilon_{i,t}$ and $\widetilde{\Phi}_{yy} = E(\widetilde{u}_{y,i}\widetilde{u}_{y,i}')$ will result in an inconsistent estimator for ρ and β unless $\gamma_x = 0$ because $E(x_i'\widetilde{\Phi}_{yy}^{-1}\widetilde{u}_{y,i}) \neq 0$ due to $E(\mu_i \widetilde{v}_{y,i}) \neq 0$. However, one can obtain a consistent LI REQML estimator for ρ and β in (1) by replacing μ_i in (1) by

$$\mu_y + \pi_y y_{i,1} + \varphi_y x_{i,1} + \psi_y \iota'(x_i - \widehat{\alpha}_x x_{i,-1} - \widehat{\beta}_x y_{i,-1}) + v_{y,i}, \tag{10}$$

where $\widehat{\alpha}_x$ and $\widehat{\beta}_x$ are preliminary consistent estimators of α_x and β_x. To see this, let

[10] We use (9) instead of $x_{i,t} = \alpha_x x_{i,t-1} + \widetilde{\beta}_x y_{i,t-1} + \widetilde{\gamma}_x \mu_i + \widetilde{\lambda}_i + \xi_{i,t}$ with $\widetilde{\beta}_x = \beta_x(1-\alpha_x)$, $\widetilde{\gamma}_x = \gamma_x(1-\alpha_x)$ and $\widetilde{\lambda}_i = (1-\alpha_x)\lambda_i$ for notational convenience.

$u_{y,i,t} = (1-\rho)v_{y,i} + \varepsilon_{i,t}$ and $\Phi_{yy} = E(u_{y,i}u'_{y,i})$. Project $\gamma_x\mu_i + \lambda_i$ on 1, $y_{i,1}$ and $x_{i,1}$ so that $\gamma_x\mu_i + \lambda_i = \widetilde{\mu}_x + \widetilde{\pi}_x y_{i,1} + \widetilde{\varphi}_x x_{i,1} + \widetilde{v}_{x,i}$ and let $\widetilde{u}_{x,i,t} = \widetilde{v}_{x,i} + \xi_{i,t}$, $\widetilde{\Phi}_{yx} = E(\widetilde{u}_{y,i}\widetilde{u}'_{x,i})$ and $\widetilde{\Phi}_{xx} = E(\widetilde{u}_{x,i}\widetilde{u}'_{x,i})$. Noting that $\widetilde{\Phi}_{yx}\widetilde{\Phi}_{xx}^{-1} \propto \iota\iota'$, one can interpret the augmented version of (1) in which μ_i is replaced by (10) as an approximation of a conditional model for $y_{i,t}$ given $y_{i,t-1}$, $x_{i,t}$, $y_{i,1}$, $x_{i,1}$ and $\widetilde{u}_{x,i}$ with an error term that is an approximation of $\widetilde{u}_{y,i} - \widetilde{\Phi}_{yx}\widetilde{\Phi}_{xx}^{-1}\widetilde{u}_{x,i}$ so that $\text{plim}_{N\to\infty} N^{-1}\sum_i(\widetilde{u}_{x,i}u'_{y,i}) = 0$. Consistency of the LI REQMLE based on (1) with μ_i replaced by (10) follows from the fact that the likelihood of $\underline{u}_{y,i} = \text{plim}_{N\to\infty} u_{y,i}$ is the same as the likelihood of y_i given $y_{i,1}$, $x_{i,1}$ and $\widetilde{u}_{x,i}$ which is equal to $\prod_{t=2}^{T} f(y_{i,t}|y_{i,t-1}, x_{i,t}, y_{i,1}, x_{i,1}, \widetilde{u}_{x,i})$. The consistency proof is similar to that given in Appendix A for the REQMLE for the panel AR(1) model without covariates. Finally, note that $\text{plim}_{N\to\infty} N^{-1}\sum_i((x_i - \widehat{\alpha}_x x_{i,-1} - \widehat{\beta}_x y_{i,-1})'\iota\iota'\Phi_{yy}^{-1}u_{y,i}) = 0$, $\text{plim}_{N\to\infty} N^{-1}\sum_i(y'_{i,-1}\Phi_{yy}^{-1}u_{y,i}) = 0$ and $\text{plim}_{N\to\infty} N^{-1}\sum_i(x'_i\Phi_{yy}^{-1}u_{y,i}) = 0$, which is in line with consistency of the REQMLE.

It is important to use consistent estimators $\widehat{\alpha}_x$ and $\widehat{\beta}_x$ in (10) instead of α_x and β_x because $\text{plim}_{N\to\infty} N^{-1}\sum_i(x'_i\iota\iota'\Phi_{yy}^{-1}u_{y,i}) \neq 0$, $\text{plim}_{N\to\infty} N^{-1}\sum_i(x'_{i,-1}\iota\iota'\Phi_{yy}^{-1}u_{y,i}) \neq 0$ and $\text{plim}_{N\to\infty} N^{-1}\sum_i(y'_{i,-1}\iota\iota'\Phi_{yy}^{-1}u_{y,i}) \neq 0$ due to $E(y_{i,s}\varepsilon_{i,t}) \neq 0$ for $s \geq t$.

One can apply GMM to (9) to obtain the preliminary consistent estimators for α_x and β_x, i.e., $\widehat{\alpha}_x$ and $\widehat{\beta}_x$. Such GMM estimators can exploit the Arellano-Bond (1991) type linear moment conditions $E(x_{i,s}(\Delta x_{i,t} - \alpha_x \Delta x_{i,t-1} - \beta_x \Delta y_{i,t-1})) = 0$, $s = 1, ..., t-2$, $t = 3, ..., T$, $E(y_{i,s}(\Delta x_{i,t} - \alpha_x \Delta x_{i,t-1} - \beta_x \Delta y_{i,t-1})) = 0$, $s = 1, ..., t-1$, $t = 3, ..., T$, and the Ahn-Schmidt (1995) type nonlinear moment conditions $E((x_{i,t} - \alpha_x x_{i,t-1} - \beta_x y_{i,t-1})(\Delta x_{i,t-1} - \alpha_x \Delta x_{i,t-2} - \beta_x \Delta y_{i,t-2})) = 0$, $t = 4, ..., T$. Alternatively, one could combine

$$y_{i,t} = \rho y_{i,t-1} + \widetilde{\beta} x_{i,t} + \widetilde{\mu}_y + \widetilde{\pi}_y y_{i,1} + \widetilde{\varphi}_y x_{i,1} + \psi_y(1-\rho)\iota'(x_i - \widehat{\alpha}_x x_{i,-1} - \widehat{\beta}_x y_{i,-1}) + u_{y,i,t}, \quad (11)$$

with a similar approximate conditional model for $x_{i,t}$, i.e.,

$$x_{i,t} = \alpha_x x_{i,t-1} + \beta_x y_{i,t-1} + \mu_x + \pi_x y_{i,1} + \varphi_x x_{i,1} + \psi_x \iota'(y_i - \widehat{\rho} y_{i,-1} - \widehat{\widetilde{\beta}} x_{i,-1}) + u_{x,i,t}, \quad (12)$$

and estimate these equations simultaneously by using the QML method while treating $\widehat{\rho}$, $\widehat{\widetilde{\beta}}$, $\widehat{\alpha}_x$ and $\widehat{\beta}_x$ as QML estimates. Although the latter approach would involve more than one equation, it can still be regarded as a LI approach as it does not fully impose the structure of the covariance matrix of the composite error vectors $\widetilde{u}_{y,i}$ and $\widetilde{u}_{x,i}$ on the system of equations (i.e., on the ψ-parameters and the parameters appearing in the covariance matrices of $u_{y,i}$ and $u_{x,i}$) unlike the FI QML approach. However, simultaneously estimating (11) and (12) may not be entirely straightforward due to non-linearities. To simplify the computations, one could instead use an iterative QML estimation procedure that alternates between the two equations and starts with consistent GMM estimates for α_x and β_x (or for ρ and β). However, this procedure is not guaranteed to converge.

When the $\xi_{i,t}$ are heteroskedastic over time, a consistent LI REQMLE can be based on (1) with μ_i replaced by (15).

A LI FEQMLE for ρ and β in (1) when $x_{i,t}$ obeys (9) with $\beta_x \neq 0$ and $\gamma_x \neq 0$ and the $\xi_{i,t}$ are homoskedastic over time can be obtained by applying the ML method to

$$\widetilde{\Delta} y_{i,t} = \rho \widetilde{\Delta} y_{i,t-1} + \widetilde{\beta} \widetilde{\Delta} x_{i,t} + \widetilde{\mu}_y + \psi_y (1-\rho)\iota'(\widetilde{\Delta} x_i - \widehat{\alpha}_x \widetilde{\Delta} x_{i,-1} - \widehat{\beta}_x \widetilde{\Delta} y_{i,-1}) + u_{y,i,t}, \quad (13)$$

where $\widetilde{\Delta} x_{i,-1} = x_{i,-1} - x_{i,1}\iota$ and $u_{y,i,t} = (1-\rho)v_{y,i} + \varepsilon_{i,t}$.

Next we discuss limited information RE QML estimation of ρ and β in the more general model (5) when, instead of (9), $x_{i,t}$, for its part, obeys the more general equation

$$x_{i,t} = \alpha_x x_{i,t-1} + \widetilde{\beta}_x y_{i,t-1} + \widetilde{\gamma}_t \mu_i + \widetilde{\zeta}_t \lambda_i + \xi_{i,t}, \quad (14)$$

where $|\alpha_x| \leq 1$, $\widetilde{\beta}_x$, $\widetilde{\gamma}_t$ and $\widetilde{\zeta}_t$ are parameters, μ_i and λ_i are independent and $\xi_{i,s}$ and $\varepsilon_{i,t}$ are independent for all s, t. A consistent LI REQMLE for ρ and β in (5) when $\widetilde{\beta}_x \neq 0$ but $\widetilde{\gamma}_t = 0$ for all t is given in Bai (2013b). Regardless of whether the $\xi_{i,t}$ are homo- or heteroskedastic, when $\widetilde{\beta}_x \neq 0$ and $\widetilde{\gamma}_t \neq 0$ for some or all t, a consistent LI REQMLE for ρ and β can be obtained by applying the ML method to (5) with μ_i replaced by

$$\mu_y + \pi_y y_{i,1} + \varphi_y x_{i,1} + \sum_{t=2}^{T} \psi_{y,t}(x_{i,t} - \widehat{\alpha}_x x_{i,t-1} - \widehat{\widetilde{\beta}}_x y_{i,t-1}) + v_{y,i}, \quad (15)$$

where $\widehat{\alpha}_x$ and $\widehat{\widetilde{\beta}}_x$ are preliminary consistent estimators of α_x and $\widetilde{\beta}_x$ such as e.g. GMM estimators due of Ahn, Lee and Schmidt (2013). The terms added to (5) ensure that $\text{plim}_{N\to\infty} N^{-1}\sum_i((\widetilde{\gamma}_s\mu_i + \widetilde{\zeta}_s\lambda_i + \xi_{i,s})u_{y,i,t}) = 0$, $\text{plim}_{N\to\infty} N^{-1}\sum_i(y_{i,1}u_{y,i,t}) = 0$ and $\text{plim}_{N\to\infty} N^{-1}\sum_i(x_{i,1}u_{y,i,t}) = 0$ for all s, t, where $u_{y,i,t} = \widetilde{\delta}_t v_{y,i} + \varepsilon_{i,t}$, and effectively consistency of the LI REQMLE for the parameters in (5), cf. the discussion on p. 8.

Finally, we discuss limited information REQML estimation of ρ and β in the model with multiple covariates and multiple time-varying individual effects, i.e.,

$$y_{i,t} = \rho y_{i,t-1} + \sum_{k=1}^{K} \beta_k (1-\rho) x_{k,i,t} + \sum_{r=1}^{R} \delta_{r,t}(1-\rho)\mu_{r,i} + \varepsilon_{i,t}, \quad (16)$$

where $\delta_{r,t}$ will be treated as parameters, and the $x_{k,i,t}$ obey

$$x_{k,i,t} = \alpha_{x,k} x_{k,i,t-1} + \widetilde{\beta}_{x,k} y_{i,t-1} + \sum_{r=1}^{R} \widetilde{\gamma}_{k,r,t} \mu_{r,i} + \widetilde{\zeta}_{k,t} \lambda_{k,i} + \xi_{k,i,t}, \tag{17}$$

where $|\alpha_{x,k}| \leq 1$, $\widetilde{\beta}_{x,k}$, $\widetilde{\gamma}_{k,r,t}$ and $\widetilde{\zeta}_{k,t}$ are parameters, $\mu_{r,i}$, $\lambda_{k,i}$ and $\lambda_{l,i}$ are independent for all k, l and r and $\xi_{k,i,s}$ and $\varepsilon_{i,t}$ are independent for all k, s, t. In this case a consistent LI REQMLE can be obtained by applying the ML method to (16) with $\mu_{r,i}$ replaced by

$$\mu_{y,r} + \pi_{y,r} y_{i,1} + \varphi_{y,r} x_{i,1} + \sum_{k=1}^{K} \sum_{t=2}^{T} \psi_{y,k,r,t} (x_{k,i,t} - \widehat{\alpha}_{x,k} x_{k,i,t-1} - \widehat{\widetilde{\beta}}_{x,k} y_{i,t-1}) + v_{y,r,i}, \tag{18}$$

where $\widehat{\alpha}_{x,k}$ and $\widehat{\widetilde{\beta}}_{x,k}$ are preliminary consistent estimators of $\alpha_{x,k}$ and $\widetilde{\beta}_{x,k}$, $k = 1, 2, ..., K$. A related FEQMLE can be obtained by replacing $y_{i,t}$ and $x_{k,i,t}$ by $y_{i,t} - y_{i,1}$ and $x_{k,i,t} - x_{k,i,1}$ for $t = 1, ..., T$ and $k = 1, 2, ..., K$ in the augmented model for y.

Remark 1: To decide whether to treat $x_{(k,)i,t}$ as strictly exogenous or predetermined w.r.t. $\varepsilon_{i,t}$ one needs to estimate (9), (14) or (17) first and then test whether $\widetilde{\beta}_{(x,)k} = 0$.

Remark 2: Applying ML to (5) with μ_i replaced by (15) and $\widehat{\widetilde{\beta}}_x = 0$ leads to the same REQMLE as applying ML to (6), irrespective of whether α_x has been estimated with $\widetilde{\beta}_x = 0$ imposed on (14) or not. A similar comment applies to the REQMLE based on (16) with $\mu_{r,i}$ replaced by (18) and $\widehat{\widetilde{\beta}}_{x,k} = 0$. On the other hand, when ($x_{i,t}$ is strictly exogenous w.r.t. $\varepsilon_{i,t}$ and) α_x has been estimated with $\beta_x = 0$ imposed on (9), the REQMLE based on (1) with μ_i replaced by (10) and $\widehat{\widetilde{\beta}}_x = 0$ is a (consistent) restricted version of the REQMLE based on (1) with μ_i replaced by (2). Similar comments apply to the FE counterparts of these REQMLEs.

Remark 3: If $\gamma_x = 0$, $\gamma_t = 0$ or $\gamma_{k,r,t} = 0$, $t = 2, ..., T$, then the term(s) involving ψ_y, $\psi_{y,t}$ or $\psi_{y,k,r,t}$, $t = 2, ..., T$, can be dropped from (10), (i.e., from (11)), (15) or (18) without causing inconsistency of the REQMLE, cf. Bai (2013b). Thus one can test weak exogeneity of a predetermined regressor by testing $\psi_y = 0$, $\psi_{y,t} = 0$ or $\psi_{y,k,r,t} = 0$, $t = 2, ..., T$, in the relevant RE model. However, even if $\gamma_x = 0$, $\gamma_t = 0$ or $\gamma_{k,r,t} = 0$, $t = 2, ..., T$, omitting the term(s) involving ψ_y, $\psi_{y,t}$ or $\psi_{y,k,r,t}$, $t = 2, ..., T$, from (13) or one of its generalizations would cause inconsistency of the FEQMLE because $\widetilde{\Delta} x_{k,i,t} - \alpha_{x,k} * \widetilde{\Delta} x_{k,i,t-1} - \beta_{x,k} \widetilde{\Delta} y_{i,t-1}$ would still contain an individual effect that is correlated with μ_i.

Remark 4: The presence of $(\widetilde{\zeta}_{(k,)t})\lambda_{(k,)i}$ in (9), (14) or (17), where $Cov(\lambda_{(k,)i}, \mu_{(r,)i}) = 0$, affects how $\widehat{\alpha}_{x(,k)}$ and $\widehat{\widetilde{\beta}}_{x(,k)}$ are computed but otherwise plays no role in the analysis.

Remark 5: the LIQMLEs discussed in this section are based on augmented models for y that include generated regressors, namely the residuals from the models for the regressors. To compute standard errors for the LIQMLEs for ρ and $\widetilde{\beta}_k$, $k = 1, ..., K$, one can still use a GLS formula but the augmented model should be viewed as an errors-in-variables model. In particular, the composite error term also includes a term like $\widetilde{\psi}_y \iota'((\widehat{\alpha}_x - \alpha_x)x_{i,-1} + (\widehat{\beta}_x - \beta_x)y_{i,-1})$ or $\sum_{r=1}^{R}(\widetilde{\delta}_{r,t} \sum_{k=1}^{K} \sum_{t=2}^{T} \psi_{y,k,r,t}((\widehat{\alpha}_{x,k} - \alpha_{x,k})x_{k,i,t-1} + (\widehat{\widetilde{\beta}}_{x,k} - \widetilde{\beta}_{x,k})y_{i,t-1}))$, which contributes to the variation in the estimates for ρ and $\widetilde{\beta}_k$, $k = 1, ..., K$. One can also use the bootstrap to compute standard errors for the LIQMLEs.

2.2 Full information QML estimators

The preceding limited information RE and FE QML estimators will no longer be consistent for ρ and β in (1) or (5) when $\xi_{i,t}$ and $\varepsilon_{i,s}$ are correlated for (some) $s < t$, in which case $x_{i,t}$ is still predetermined with respect to $\varepsilon_{i,t}$, or when $\xi_{i,t}$ and $\varepsilon_{i,s}$ are contemporaneously correlated, in which case $x_{i,t}$ is endogenous even if $\gamma_x = 0$ in (9) or $\gamma_t = 0$, $t = 2, ..., T$ in (14). Note that in both cases $x_{i,t}$ is also still affected by lags of $\varepsilon_{i,t}$ through $y_{i,t-1}$. In both cases we can adopt a Full Information QML approach to estimation that is based on a VAR model. Upon substituting the RHS of (14) for $x_{i,t}$ in (5) and letting $\xi_{i,t}$ absorb $\widetilde{\zeta}_t \lambda_i$, we obtain the following VAR model:

$$\begin{bmatrix} y_{i,t} \\ x_{i,t} \end{bmatrix} = \begin{bmatrix} \rho + \widetilde{\beta}_x \widetilde{\beta} & \alpha_x \widetilde{\beta} \\ \widetilde{\beta}_x & \alpha_x \end{bmatrix} \begin{bmatrix} y_{i,t-1} \\ x_{i,t-1} \end{bmatrix} + \begin{bmatrix} (\widetilde{\delta}_t + \widetilde{\gamma}_t \widetilde{\beta}) \\ \widetilde{\gamma}_t \end{bmatrix} \mu_i + \begin{bmatrix} \varepsilon_{i,t} + \widetilde{\beta} \xi_{i,t} \\ \xi_{i,t} \end{bmatrix} \tag{19}$$

with $\widetilde{\beta}_x = \beta_x(1 - \alpha_x)$, $\widetilde{\beta} = \beta(1 - \rho)$, $\widetilde{\delta}_t = \delta_t(1 - \rho)$ and $\widetilde{\gamma}_t = \gamma_t(1 - \alpha_x)$. This model can be written more succinctly as

$$z_{i,t} = A z_{i,t-1} + \widetilde{\zeta}_t \mu_i + \omega_{i,t}, \tag{20}$$

where $z_{i,t} = (y_{i,t},\ x_{i,t})'$ and the other symbols are defined implicitly. From this point onwards we will focus the discussion on the case where $\xi_{i,t}$ and $\varepsilon_{i,s}$ are only contemporaneously correlated (i.e., when $s = t$). To obtain the FI REQMLE for ρ and $\widetilde{\beta}$ (or β) in (20), apply the ML method to the model with μ_i replaced by

$$\mu + \phi z_{i,1} + v_i = \mu + \pi y_{i,1} + \varphi x_{i,1} + v_i,$$

that is, to

$$z_{i,t} = Az_{i,t-1} + \widetilde{\zeta}_t(\mu + \pi y_{i,1} + \varphi x_{i,1}) + u_{i,t}, \quad (21)$$

where $u_{i,t} = \widetilde{\zeta}_t v_i + \omega_{i,t}$.[11] To achieve identification we can impose $\sigma_v^2 = 1$. Note that not only the $\widetilde{\zeta}_t$ but also $\widetilde{\beta}$ appears in both the mean equation and the covariance matrix of the $u_{i,t}$. However, if $Cov(\xi_{i,t}, \varepsilon_{i,t}) \neq 0$ for all t, then $\widetilde{\beta}$ is only identified by the mean equations and the FIQMLE for ρ and $\widetilde{\beta}$ can be computed more easily by leaving the covariance matrices for the $\omega_{i,t}$ unrestricted without affecting its efficiency. On the other hand, if $Cov(\xi_{i,t}, \varepsilon_{i,t}) = 0$ for at least some t, then our FI QMLE for ρ and $\widetilde{\beta}$ (or β) is more efficient than (a version of) the FI QMLE discussed on p.11 in Bai (2013b) as the latter does not exploit the implied constraint(s) on the estimation problem, i.e., on the covariance matrix of the errors. If $Cov(\xi_{i,t}, \varepsilon_{i,t}) = 0$ for at least some t, then it is more practical to apply the FI QMLE to the system that consists of (5) and (14) with μ_i replaced by $\mu + \pi y_{i,1} + \varphi x_{i,1} + v_i$ rather than to (21).

Bai (2013b) discusses how to implement the ECM algorithm of Meng and Rubin (1993) for computing (FI) QMLEs based on likelihood functions very similar to (21) apart from the fact that they do not contain a slope parameter such as $\widetilde{\beta}$ that appears in both the mean equation and the covariance matrix; only the $\widetilde{\zeta}_t$ appear in both the mean and the variance. His algorithm can be applied directly to estimation of (21) when $Cov(\xi_{i,t}, \varepsilon_{i,t}) \neq 0$ for all t or when $Cov(\xi_{i,t}, \varepsilon_{i,t}) = 0$ for at least some t *and* the $\widetilde{\beta}$ that appears in the covariance matrix of the $u_{i,t}$ is replaced by a different parameter, say β^F, thereby removing a constraint. The latter approach would result in an estimator that is less efficient than the FI REQMLE. In the latter case, the FI REQMLE can be obtained by applying the algorithm to (5) and (14) with μ_i replaced by $\mu + \pi y_{i,1} + \varphi x_{i,1} + v_i$.

Consistency of the FI REQMLE for ρ and $\widetilde{\beta}$ (and α_x and $\widetilde{\beta}_x$) in (21) can be shown similarly to the LI REQMLE for the panel AR(1) model with a time-varying individual effect, cf. Bai (2013b). Note that the equation for $x_{i,t}$ is included in system (21) to achieve consistency rather than efficiency; only estimating the first equation in system (21) would result in an inconsistent estimator of the parameters unless $\widetilde{\gamma}_t = 0$, $t = 2, ..., T$.

[11] When $\xi_{i,t}$ and $\varepsilon_{i,s}$ are correlated for (some) $s < t$, then the $\omega_{i,t}$ will be correlated with some lag(s) of $y_{i,t}$ and possibly $x_{i,t}$. In this case we can still obtain a consistent FIQMLE by including additional terms in the model, cf. the approach in Blundell and Smith (1991). For instance, if $\xi_{i,t}$ is correlated with $\varepsilon_{i,t-1}$ and $\varepsilon_{i,t-2}$, then add the term $\widetilde{\beta}\widetilde{\tau}_3 y_{i,1}$ to the equation for $y_{i,3}$, the term $\widetilde{\tau}_3 y_{i,1}$ to the equation for $x_{i,3}$, the term $\widetilde{\beta}(\widetilde{\tau}_2 y_{i,1} + \widetilde{\vartheta}_2 x_{i,1})$ to the equation for $y_{i,2}$ and the term $\widetilde{\tau}_2 y_{i,1} + \widetilde{\vartheta}_2 x_{i,1}$ to the equation for $x_{i,2}$.

Instead of a QML estimator, we can also use a GMM estimator to estimate ρ and $\widetilde{\beta}$ (and α_x and $\widetilde{\beta}_x$). Let $z_{i,-1} = (z'_{i,1} \ldots z'_{i,T-1})'$, $u_i = (u'_{i,2} \ldots u'_{i,T})'$, $\omega_i = (\omega'_{i,2} \ldots \omega'_{i,T})'$, $\overline{\zeta} = (\widetilde{\zeta}'_2, ..., \widetilde{\zeta}'_T)'$ and $\Phi = E(u_i u'_i) = \sigma^2_v \overline{\zeta}\overline{\zeta}' + \Psi$ with $\sigma^2_v = 1$ and $\Psi = E(\omega_i \omega'_i) = diag(\Omega_2, ..., \Omega_T) = \Psi(\beta^F, \Sigma_2, ..., \Sigma_T)$, where $\Omega_t = E(\omega_{i,t}\omega'_{i,t}) = \Omega_t(\beta^F, \Sigma_t)$ and $\Sigma_t = E((\varepsilon_{i,t}, \xi_{i,t})'(\varepsilon_{i,t}, \xi_{i,t})) = \Sigma_t((\Sigma_t)_{j,k})$, $t = 2, ..., T$, with $j, k \in \{1, 2\}$ and $j \leq k$, and where β^F is an unconstrained parameter taking the place of $\widetilde{\beta}$ in Ψ. Also, let $S_k = I_T \otimes J_k$, $k = 1, 2, 3, 4$, where the J_k are four different 2×2 matrices each with one non-zero element that is equal to one. Then a related GMM estimator exploits $E(z'_{i,-1} S_k \Phi^{-1} u_i) = 0$, $k = 1, 2, 3, 4$, $E(\overline{\zeta}' \Phi^{-1} u_i) = 0$, $E(z_{i,1} \overline{\zeta}' \Phi^{-1} u_i) = 0$, $E(tr((u_i u'_i - \Phi)\partial\Phi^{-1}/\partial(\Sigma_t)_{j,k})) = 0$ for $t = 2, ..., T$, $j = 1, 2$ and $k = 1, 2$ with $j \leq k$, $E(tr((u_i u'_i - \Phi)\partial\Phi^{-1}/\partial\beta^F)) = 0$ only when $(\Sigma_t)_{1,2} = Cov(\xi_{i,t}, \varepsilon_{i,t}) = 0$ for some or all t, and $E((\mu + \phi z_{i,1}) e'_t \Phi^{-1} u_i) - \frac{1}{2} E(tr((u_i u'_i - \Phi)\partial\Phi^{-1}/\partial\overline{\zeta}_t)) = 0$, $t = 1, ..., 2(T-1)$, where e_t is the $t - th$ column of the identity matrix $I_{2(T-1)}$ as well as the parameter restrictions on A and Ψ. The optimal version of this GMM estimator is efficient when $Cov(\xi_{i,t}, \varepsilon_{i,t}) \neq 0$ for all t. If $Cov(\xi_{i,t}, \varepsilon_{i,t}) = 0$ for some or all t, then an optimal GMM estimator that exhausts the moment conditions that are directly based on the system that consists of (5) and (14) with μ_i replaced by $\mu + \pi y_{i,1} + \varphi x_{i,1} + v_i$ is more efficient.

Remark 6: The number of initial observations of y and x that should be included in the augmented equation(s) depends on the lag structures, that is, on the lag lengths of the original equations for $y_{i,t}$ and $x_{i,t}$.

Remark 7: The standard errors for the FIQMLEs can be computed in the standard way unlike those for some of the LIQMLEs, cf. remark 5.

Remark 8: The FIQML approach can, of course, also be used when $\xi_{i,t}$ and $\varepsilon_{i,s}$ are uncorrelated for all $s, t \in \{2, ..., T\}$ and offers an alternative to the LIQMLEs that have been discussed earlier. When the data are i.i.d. across the individuals and Gaussian, both the FIMLEs and the LIMLEs for ρ and $\widetilde{\beta}$ are asymptotically efficient. However, the LIQMLE and FIQMLE have different finite samples properties. Furthermore, the LI approach may be more attractive than the FI approach from a computational point of view especially when the number of regressors is large. Finally, note that even if $\xi_{i,t}$ and $\varepsilon_{i,s}$ are uncorrelated for all $s, t \in \{2, ..., T\}$, the $\Omega_t = E(\omega_{i,t}\omega'_{i,t})$ related to the model in (21) are still non-diagonal unless $\widetilde{\beta} = 0$.

Remark 9: To test whether $y_{i,t}$ is affected by $x_{i,t-1}$ rather than by $x_{i,t}$ one might consider estimating an unrestricted version of the VAR(1) model in (21) and testing whether $(\Omega_s)_{1,2} = (\Sigma_s)_{1,2}$ for some (or all) $s \in \{2,, ..., T\}$. To implement such a test one would require the value(s) of (the) $(\Sigma_s)_{1,2}$, e.g. $(\Sigma_s)_{1,2} = 0$, which, however, is/are usually unknown. This suggests that the VAR based approach does not allow one to distinguish between a structural model where $y_{i,t}$ is affected by $x_{i,t}$ but not by $x_{i,t-1}$ and a structural model where $y_{i,t}$ is affected by $x_{i,t-1}$ but not by $x_{i,t}$ without making an assumption about the value of at least one of the covariances $(\Sigma_s)_{1,2}$, $s = 2, ..., T$. One can solve this identification problem by following a two step testing procedure. In the first step one estimates the parameters of the two competing models. In the second step one first estimates an equation for $y_{i,t}$ that includes a convex combination of the two estimated equations for $y_{i,t}$ from step 1, e.g. $y_{i,t} = \lambda(\widehat{\rho} y_{i,t-1} + \widehat{\widehat{\beta}} x_{i,t}) + (1-\lambda)(\widehat{\rho}_{(t-1)} y_{i,t-1} + \widehat{\widehat{\beta}}_{(t-1)} x_{i,t-1}) + \widetilde{\delta}_t(\mu + \pi y_{i,1} + \varphi x_{i,1} + \psi_y \iota'(x_i - \widehat{\alpha}_x x_{i,-1} - \widehat{\beta}_x y_{i,-1})) + u_{i,t}$, where $\widehat{\rho}_{(t-1)}$ and $\widehat{\widehat{\beta}}_{(t-1)}$ are estimators for ρ and β in the model in which $y_{i,t}$ is affected by $x_{i,t-1}$ rather than by $x_{i,t}$, $\widetilde{\delta}_t = \delta_t(1-\rho)$ and $u_{i,t} = \widetilde{\delta}_t v_i + \eta_{i,t}$ with v_i and $\eta_{i,t}$ error terms, and then tests whether $y_{i,t}$ is affected by $x_{i,t-1}$ rather than by $x_{i,t}$ by testing $\lambda = 0$.

Remark 10: Once the correct lag structures of x and y in the equation for $y_{i,t}$ have been determined, one can proceed to test whether $(\Sigma_s)_{1,2} = 0$, $s = 2, ..., T$ using a FI approach. If this hypothesis is not rejected, then one may wish to consider using a LI estimator, see also remark 8.

The model in (20) can also be estimated by a FI FEQMLE. In this case we estimate the system

$$\widetilde{\Delta} z_{i,t} = A \widetilde{\Delta} z_{i,t-1} + \widetilde{\lambda} + \widetilde{\lambda}_i + \widetilde{\zeta}_t \mu_i + \omega_{i,t}, \tag{22}$$

where $\widetilde{\Delta} z_{i,t} = z_{i,t} - z_{i,1}$, $\widetilde{\lambda} = (\widetilde{\lambda}_1\ \widetilde{\lambda}_2)'$ and $\widetilde{\lambda}_i = (\widetilde{\lambda}_{1,i}\ \widetilde{\lambda}_{2,i})'$. The QMLE based on (22) is consistent, cf. the FI *RE*QMLE in Bai (2013b); in this case there is no need to project the three individual effects, namely $\widetilde{\lambda}_{1,i}$, $\widetilde{\lambda}_{2,i}$ and μ_i, on some lags of $\widetilde{\Delta} z_{i,t}$. Indeed $\widetilde{\Delta} z_{i,1} = 0$. Note that the number of individual effects in a FE model like (22) grows with the number of equations in the system, i.e., the number of regressors in the model. To identify the model we can impose $\sigma^2_\mu = 1$.

To derive (22), we have first rewritten (20) as

$$\widetilde{\Delta} z_{i,t} = A \widetilde{\Delta} z_{i,t-1} - (I - A) z_{i,1} + \widetilde{\zeta}_t \mu_i + \omega_{i,t}. \tag{23}$$

If $Cov(\xi_{i,t}, \varepsilon_{i,t}) = 0$ for some or all t, then it is again more practical to apply the FI

QMLE to a transformed version of the original system, which consists of

$$\begin{aligned} \widetilde{\Delta} y_{i,t} &= \rho \widetilde{\Delta} y_{i,t-1} + \widetilde{\beta} \widetilde{\Delta} x_{i,t} + \widetilde{\lambda}_3 + \widetilde{\lambda}_{3,i} + \widetilde{\delta}_t \mu_i + \varepsilon_{i,t} \text{ and} \\ \widetilde{\Delta} x_{i,t} &= \alpha_x \widetilde{\Delta} x_{i,t-1} + \beta_x (1-\alpha_x) \widetilde{\Delta} y_{i,t-1} + \widetilde{\lambda}_2 + \widetilde{\lambda}_{2,i} + \widetilde{\gamma}_t \mu_i + \xi_{i,t}. \end{aligned} \tag{24}$$

Note that, similar to the RE case, the equation for $\widetilde{\Delta} x_{i,t}$ is included in system (22) or (24) to achieve consistency rather than efficiency; only estimating the first equation in system (22) or (24) would result in an inconsistent estimator of the parameters therein unless $\widetilde{\gamma}_t = 0$, $t = 2, ..., T$.

3 Random Effects or Fixed Effects?

To choose between a RE and FE approach, one can use a Hausman test. However, for the FEMLEs discussed above there are usually several RE counterparts. For instance, one can compare the FEMLE for model (1) with an REMLE that allows for correlation between μ_i and $(x_{i,1}\ x_i')'$ or with an REMLE that does not. When there are several RE counterparts for a particular FEMLE, one may wish to compare this FE estimator with the REMLE for the most general model for which the FEMLE is still consistent as we will see below. In this section we will outline Hausman tests for LI MLEs for models with strictly exogenous regressors similar to (1) and (5) in section 2.1.1. and for FI MLEs for a system with lagged predetermined regressors similar to (20) with $E(\omega_{i,t}\omega_{i,t}') = E((\varepsilon_{i,t}, \xi_{i,t})'(\varepsilon_{i,t}, \xi_{i,t})) = diag(\sigma_\varepsilon^2, \sigma_\xi^2)$, $t = 2, ..., T$.

An REMLE for (1) that is comparable to the FEMLE based on (3) is based on

$$y_{i,t} = \rho y_{i,t-1} + \widetilde{\beta} x_{i,t} + \widetilde{\mu} + \widetilde{\pi} y_{i,1} + \{\widetilde{\varphi} x_{i,1} + \widetilde{\gamma}' x_i\} + u_{i,t}, \tag{25}$$

where $u_{i,t} = (1-\rho)v_i + \varepsilon_{i,t}$. Note that unlike the REMLE, the FEMLE for (1) always allows for correlation between μ_i and $(x_{i,1}\ x_i')'$. The Hausman test-statistic takes the familiar form:

$$H = (\widehat{\theta}_{FE} - \widehat{\theta}_{RE})'(Avar(\widehat{\theta}_{FE}) - Avar(\widehat{\theta}_{RE}))^{-1}(\widehat{\theta}_{FE} - \widehat{\theta}_{RE}), \tag{26}$$

where θ is a parameter vector and $H \sim \chi^2(\dim(\theta))$ when both $\widehat{\theta}_{RE}$ and $\widehat{\theta}_{FE}$ are consistent and $\widehat{\theta}_{RE}$ is efficient.

The MLE based on (7) is not only the FEMLE for (5) but also the FEMLE for a slightly more general model that also includes a time-invariant individual effect:

$$y_{i,t} = \rho y_{i,t-1} + \beta(1-\rho)x_{i,t} + (1-\rho)\mu_{1,i} + \delta_t(1-\rho)\mu_{2,i} + \varepsilon_{i,t}. \tag{27}$$

Therefore, one can compare the FEMLE based on (7) with any REMLE that is based on a version of

$$\begin{aligned} y_{i,t} \quad = \quad & \rho y_{i,t-1} + \widetilde{\beta} x_{i,t} + \{\widetilde{\mu}_1 + \widetilde{\pi}_1 y_{i,1}\} + \{\widetilde{\varphi}_1 x_{i,1} + \widetilde{\gamma}_1' x_i\} + \\ & \widetilde{\delta}_t(\mu_2 + \pi_2 y_{i,1} + \{\varphi_2 x_{i,1} + \gamma_2' x_i\}) + u_{i,t}, \end{aligned} \tag{28}$$

where $u_{i,t} = \{(1-\rho)v_{1,i}\} + \widetilde{\delta}_t v_i + \varepsilon_{i,t}$. Note that it is possible that a REMLE that is based on a more complete version of (28) is consistent whereas a REMLE that is based on a more parsimonious version of (28) is inconsistent leading to a different outcome for the Hausman test and potentially less efficient estimation and inference.

Similar to the previous example, the FI MLE based on (24) is not only the FI FEMLE for the system that consists of (5) and (14) but also the FI FEMLE for a slightly more general system that consists of

$$y_{i,t} = \rho y_{i,t-1} + \beta(1-\rho)x_{i,t} + (1-\rho)\mu_{1,i} + \delta_t(1-\rho)\mu_{2,i} + \varepsilon_{i,t} \tag{29}$$

and

$$x_{i,t} = \alpha_x x_{i,t-1} + \beta_x(1-\alpha_x)y_{i,t-1} + (1-\alpha_x)\mu_{3,i} + \gamma_t(1-\alpha_x)\mu_{2,i} + \zeta_t(1-\alpha_x)\lambda_i + \xi_{i,t}, \tag{30}$$

where $Cov(\mu_{1,i}, \mu_{3,i}) \neq 0$ is possible. Thus one can compare the FI FEMLE with an FI REMLE that assumes $\mu_{1,i} = \mu_{3,i} = 0$ for all $i \in \{1, ..., N\}$ or with an FI REMLE that allows for $\mu_{1,i} \neq 0$ and $\mu_{3,i} \neq 0$ for all $i \in \{1, ..., N\}$.

4 The finite sample performance of the estimators

In this section we compare through Monte Carlo simulations the finite sample properties of several LI and FI REMLEs and GMM estimators for the slope coefficients in an equation from various bivariate panel VAR(1) models with time-invariant individual effects and some related t-tests. In all our simulation experiments the regressors are correlated with the time-invariant error components of the equations and the idiosyncratic errors are contemporaneously uncorrelated. We study how the properties of the estimators are affected if we change (1) the joint distributions of the initial conditions or (2) the correlation between the time-invariant error components of the equations or (3) the coefficients of the regressors. The time series for $\{(y_{i,t}\ x_{i,t})\}$ were generated according to

$$\begin{pmatrix} y_{i,t}-\mu_{y,i} \\ x_{i,t}-\mu_{x,i} \end{pmatrix} = \begin{pmatrix} \rho & \widetilde{\beta} \\ \widetilde{\beta}_x & \alpha_x \end{pmatrix} \begin{pmatrix} y_{i,t-1}-\mu_{y,i} \\ x_{i,t-1}-\mu_{x,i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i,t} \\ \xi_{i,t} \end{pmatrix}$$

We conducted the simulation experiments for $(T, N) = (4, 100)$, $(9, 100)$, $(4, 500)$ or $(9, 500)$ and two sets of slope coefficients: A) $\rho = \alpha_x = 0.2$ and $\widetilde{\beta} = \widetilde{\beta}_x = 0.6$ and B) $\rho = \alpha_x = 0.6$ and $\widetilde{\beta} = \widetilde{\beta}_x = 0.2$. In all simulation experiments the vectors of the error components $(\mu_{y,i}\ \mu_{x,i})'$ and $(\varepsilon_{i,t}\ \xi_{i,t})'$ are mutually i.i.d. across i and t. We assumed that $(\varepsilon_{i,t}\ \xi_{i,t})' \sim i.i.d.\mathrm{N}(0, diag(\sigma^2_\varepsilon, \sigma^2_\xi)$ with $\sigma^2_\varepsilon = \sigma^2_\xi = 1$ for all $i \in \{1, ..., N\}$ and $t \in \{2, ..., T\}$. Given the values of N and T and ρ, α_x, $\widetilde{\beta}$ and $\widetilde{\beta}_x$ we considered three designs of the experiments:

I) $\{(y_{i,t}, x_{i,t})\}$ is covariance stationary and $\mu_{y,i} = \mu_{x,i} \sim i.i.d.\mathrm{N}(0, \sigma^2_\mu)$ with $\sigma^2_\mu = 1$.

II) $y_{i,1} = \mu_{y,i} + \varepsilon_{i,1}$, $x_{i,1} = \mu_{x,i} + \xi_{i,1}$ and $\mu_{y,i} = \mu_{x,i} \sim i.i.d.\mathrm{N}(0, \sigma^2_\mu)$ with $\sigma^2_\mu = 1$ and $(\varepsilon_{i,1}\ \xi_{i,1})' \sim i.i.d.\mathrm{N}(0, diag(\sigma^2_\varepsilon, \sigma^2_\xi)$.

III) $y_{i,1} = \mu_{y,i}+\varepsilon_{i,1}$, $x_{i,1} = \mu_{x,i}+\xi_{i,1}$, $\mu_{y,i} = 0.8*\mu_{z,i}+0.6*\lambda_{y,i}$ and $\mu_{x,i} = 0.8*\mu_{z,i}+0.6*\lambda_{x,i}$ with $(\mu_{z,i}\ \lambda_{y,i}\ \lambda_{x,i})' \sim i.i.d.$ $\mathrm{N}(0, diag(1, 1, 1)$ and $(\varepsilon_{i,1}\ \xi_{i,1})' \sim i.i.d.\mathrm{N}(0, diag(\sigma^2_\varepsilon, \sigma^2_\xi)$.

Note that in experiments II and III $\{(y_{i,t}\ x_{i,t})\}$ is mean stationary but not covariance stationary.

We considered the following estimators for ρ and $\widetilde{\beta}$: a single equation two-step optimal Arellano-Bond (AB) type GMM estimator based on the linear moment conditions given in section 2.1.2; two versions of a single equation multi-step optimal Ahn-Schmidt (AS) type GMM estimator as described in section 2.1.2, i.e., the original non-linear three-step version based on numerical optimization, OPAS, and a linearized four-step version using preliminary two-step optimal AB estimates in the differenced part of the non-linear moment conditions, ABAS; three versions of the LI REMLE, i.e., an infeasible version replacing $\widehat{\alpha}_x$ and $\widehat{\widetilde{\beta}}_x$ by the true values of α_x and $\widetilde{\beta}_x$, INFELIML, and two versions based on the aforementioned two versions of the single equation AS GMM estimators for α_x and $\widetilde{\beta}_x$, OPASLIML and ABASLIML, and finally the FI REMLE.

In the experiments we imposed homoskedasticity on the LI and FI likelihood functions and to ensure that the estimates of the covariance matrices $E(u_i u_i')$ were positive definite (PD), where $u_i = \widetilde{v}_{y,i}\iota + \varepsilon_i$ or $u_i' = (\widetilde{v}_{y,i}\iota + \varepsilon_i \; \widetilde{v}_{x,i}\iota + \xi_i)'$ contains the composite errors of the augmented model equation(s), we also imposed the restrictions $\sigma_\varepsilon^2 > 0$ and $Var(\widetilde{v}_{y,i}) > 0$ in the LI case and $\sigma_\varepsilon^2 > 0$, $\sigma_\xi^2 > 0$, $\sigma_\xi^2 + (T-1)Var(\widetilde{v}_{x,i}) > 0$ and $(\sigma_\xi^2 + (T-1)Var(\widetilde{v}_{x,i}))(\sigma_\varepsilon^2 + (T-1)Var(\widetilde{v}_{y,i})) - (T-1)^2(Cov(\widetilde{v}_{x,i}, \widetilde{v}_{y,i}))^2 > 0$ in the FI case. The restrictions in the LI case are stronger than those in the FI case and could be relaxed to $\sigma_\varepsilon^2 > 0$ and $\sigma_\varepsilon^2 + (T-1)Var(\widetilde{v}_{y,i}) > 0$. This would result in LIQMLEs with different, probably worse finite sample properties, cf. Bun et al. (2017). On the other hand, we could also strengthen the restrictions on the parameters in the FI case by adding $Var(\widetilde{v}_{y,i}) > 0$ and $Var(\widetilde{v}_{x,i}) > 0$. However, this would render the inclusion of a nonlinear inequality restriction unavoidable.[12] The AS GMM estimators did not exploit homoskedasticity.

Finally, we allowed for time effects by subtracting cross-sectional averages from the data.

For the estimators we calculated the bias and the MSE and in some cases the average standard error (s.e.). The s.e. of the AB estimator is based on Windmeijer's (2005) formula.

We also computed the empirical size, i.e., rejection frequency (rej.f.) of Wald tests based on the AB estimator, a LIMLE and the FIMLE. All tests had a nominal size of 5%. The relevant simulations results only give an indication of the differences in the size properties of such Wald tests because in the LIML based test we used the infeasible LIMLE which allowed us to use conventional estimators for the standard errors, cf. remark 5 above. Note also that in practice one should use robust inference procedures corresponding to these estimation methods, cf. Kruiniger (2016).

The simulation results are reported in six tables which differ with respect to the

[12] Note also that in the FI case imposing restrictions on the parameters to ensure that $E(u_i u_i')$ is PD leads to an increasingly complicated constrained maximization problem when the dimension of the system of equations increases and even more so in case one allows for heteroskedasticity over time. One can avoid imposing restrictions by using Bai's (2013b) ECM algorithm, which produces estimates that satisfy them. However, these ECM estimators will have different, probably worse finite sample properties than the constrained (FI)QMLEs, cf. Bun et al. (2017). Note that the latter may produce estimates that are on the boundary of the parameter space and correspond to higher likelihood values than the ECM estimates.

designs of the experiments (I, II or III) and the values of the slope coefficients (A or B). Inspection of the results leads to the following conclusions:

1. The LI MLEs have better finite sample properties, i.e., smaller bias and MSE than the corresponding GMM estimators.
2. The LI MLEs are often more precise, i.e., have smaller MSE than the FI MLEs.
3. When the time-dimension T is not small relative to the cross-section dimension N, the (LI and FI) MLE based Wald tests have much better size properties than GMM based Wald tests.

5 Concluding remarks

In this paper we discussed large N, fixed T consistent limited and full information RE and FE Quasi ML and GMM estimators for variations of the conditional panel AR(1) model for $y_{i,t}$ with one additional regressor $x_{i,t}$ under alternative assumptions about the nature of that regressor. We found that LI QMLEs that are based on a single augmented equation can still be derived when the equations of the dependent variable and the regressor depend on the same factor as long as the idiosyncratic errors of the equations for $y_{i,t}$ and $x_{i,s}$ are uncorrelated. On the other hand, if these idiosyncratic errors are correlated, then an FI (Q)ML estimator based on a system of augmented equations is required.

When ρ in (1) or (5) equals unity, the $x_{i,t}$ and μ_i drop out of the model and ρ is first-order underidentified under time series homoskedasticity, cf. Kruiniger (2013). It follows that close to this point in the parameter space, the QMLEs for ρ have slower rate of convergence and non-standard distributions. However, when $\rho = 1$ the QMLEs for $\widetilde{\beta} = \beta(1-\rho)$ still converges at the standard rate ($\sqrt{N}$) to a normal random variable due to block-diagonality of the limiting Hessian of the log-likelihood (scaled by N). Kruiniger (2016) discusses Quasi LM tests for hypotheses about ρ that have asymptotically correct size in a uniform sense.

Finally, we expect that the proposed ML estimators for dynamic panel data models with additional regressor(s) are still consistent in a large N, T setting as they are extensions of estimators for similar models without additional regressors for which large N, T consistency has been proven in Bai (2013a and 2013b).

A Proof of Theorem 2 in Kruiniger, 2013 (Consistency of FE-/REQMLE for a panel AR(1) model):

We first prove consistency of the FEQMLE for ρ by verifying the conditions of Theorem 2.1 in Newey and McFadden (1994, henceforth NMcF).[13] The FEQMLE for ρ is based on the quasi likelihood function corresponding to the following auxiliary model

$$y_i - y_{i,1}\iota = \rho(y_{i,-1} - y_{i,1}\iota) + u_i, \tag{31}$$

where $-1 < \rho \leq 1$ and $u_i = \widetilde{v}_{i,1}\iota + \varepsilon_i \sim N(0, \Phi)$ with $\widetilde{v}_{i,1} = (\rho - 1)v_{i,1}$, $v_{i,1} = y_{i,1} - \mu_i$ and $\Phi = \Phi(\varphi) = \widetilde{\sigma}_v^2 \iota\iota' + \Psi$, where φ is the vector comprising all (co-)variance parameters.

The Basic Assumptions imply that $\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N \varepsilon_i\varepsilon_i' = \Psi = diag(\overline{\sigma}_2^2, ..., \overline{\sigma}_T^2)$, $\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N \widetilde{v}_{i,1}^2 = \widetilde{\sigma}_v^2 < \infty$ and $\text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^N \widetilde{v}_{i,1}\varepsilon_{i,t} = 0$ for $t = 2, ..., T$, where $\overline{\sigma}_s^2 = \lim_{N\to\infty} N^{-1} \sum_{i=1}^N \sigma_{i,s}^2$ with $\sigma_{i,s}^2 = E(\varepsilon_{i,s}^2)$ for $s = 2, ..., T$.

The quasi log-likelihood for the conditional panel AR(1) model with FE is given by

$$l_{FE}(r, f) = -\frac{1}{2}N(T-1)\ln 2\pi - \frac{N}{2}\ln|F| - \frac{1}{2}\sum_{i=1}^N (\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1})' F^{-1} (\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1}), \tag{32}$$

where $F = F(f)$. We can express $\widetilde{\Delta}y_{i,-1}$ in terms of $v_{i,1}$ and ε_i

$$\widetilde{\Delta}y_{i,-1} = Pu_i = P\iota(\rho - 1)v_{i,1} + P\varepsilon_i, \tag{33}$$

where

$$P = P(\rho) = \begin{pmatrix} 0 & . & . & 0 & 0 & 0 \\ 1 & 0 & & & 0 & 0 \\ \rho & 1 & 0 & & & 0 \\ . & \rho & 1 & 0 & & . \\ . & & \rho & 1 & 0 & . \\ \rho^{T-3} & . & . & \rho & 1 & 0 \end{pmatrix}. \tag{34}$$

Next, we can rewrite (32) using that

$$\begin{aligned} &\widetilde{\Delta}y_i - r\widetilde{\Delta}y_{i,-1} = (\rho - r)\widetilde{\Delta}y_{i,-1} + u_i = ((\rho - r)P + I)u_i = \\ &((\rho - r)P\iota + \iota)(\rho - 1)v_{i,1} + ((\rho - r)P + I)\varepsilon_i. \end{aligned} \tag{35}$$

[13] The proof assumes that the parameter space is compact. If one does not like this assumption, then one can prove consistency of the QMLE for ρ by using a version of Theorem 2.7 in NMcF after a reparametrization so that the likelihood function is concave, see NMcF for details.

It follows from (35), our assumptions and the Markov Law of Large Numbers that $N^{-1}l_{FE}(r,f)$ converges uniformly in probability to a nonrandom function, $\bar{l}_{FE}(r,f)$ say. $N^{-1}l_{FE}(r,f)$ would converge uniformly in probability to exactly the same function $\bar{l}_{FE}(r,f)$ if the $\widetilde{v}_{i,1}=(\rho-1)v_{i,1}$ and ε_i were i.i.d. and normal with $E(\widetilde{v}_{i,1}^2)=\widetilde{\sigma}_v^2$, $E(\widetilde{v}_{i,1}\varepsilon_i)=0$, and $E(\varepsilon_i\varepsilon_i')=\Psi$. Therefore to verify the other conditions of Theorem 2.1 of NMcF we can use Theorem 2.5 of NMcF.

To show that ρ and φ are uniquely identified when $-1<\rho\leq 1$ we proceed as follows: Let $g(\widetilde{\Delta} y_i|\rho,\varphi)$ be the normal pdf of $\widetilde{\Delta} y_i$.

From (35) we obtain $(\widetilde{\Delta} y_i - r\widetilde{\Delta} y_{i,-1})'F^{-1}(\widetilde{\Delta} y_i - r\widetilde{\Delta} y_{i,-1}) = u_i'((\rho-r)P+I)'F^{-1}\times ((\rho-r)P+I)u_i$.

Note that $F=[f_{i,j}]=F(f)$ is PD as long as $\overline{s}_t^2>0$ for some $t\geq 2$, and that $((\rho-r)P+I)$ is nonsingular for any $-1<\rho,r\leq 1$. Hence $((\rho-r)P+I)'F^{-1}((\rho-r)P+I)$ is PD.

Furthermore, given the specific structure of F and P, $((\rho-r)P+I)'F^{-1}((\rho-r)P+I)=\Phi^{-1}$ if and only if $f=\varphi$ and $r=\rho$, unless $\rho=1$ and either $T<4$ or $\overline{\sigma}_s^2/\overline{\sigma}_{s-1}^2=\overline{\sigma}_3^2/\overline{\sigma}_2^2\neq 1$ for all $s\in\{3,...,T-1\}$. If $\rho=1$, $T\geq 4$ and $\overline{\sigma}_s^2/\overline{\sigma}_{s-1}^2=\overline{\sigma}_3^2/\overline{\sigma}_2^2\neq 1$ for all $s\in\{3,...,T-1\}$, both $((\rho-r)P+I)'F^{-1}((\rho-r)P+I)=\Phi^{-1}$ and $\det(F)=\det(\Phi)$ not only hold for $r=1$, $f_{i,i}=\overline{\sigma}_{i+1}^2$, $i=1,...,T-1$, $f_{i,j}=0$, $i\neq j$ but also for $r=\overline{\sigma}_3^2/\overline{\sigma}_2^2$, $f_{i,i}=\overline{\sigma}_2^2-\overline{\sigma}_3^2+\overline{\sigma}_2^2(\overline{\sigma}_3^2/\overline{\sigma}_2^2)^i$, $i=1,...,T-2$, $f_{T-1,T-1}=\overline{\sigma}_2^2-\overline{\sigma}_3^2+\overline{\sigma}_2^2(\overline{\sigma}_3^2/\overline{\sigma}_2^2)^{T-2}((\overline{\sigma}_3^2/\overline{\sigma}_2^2)-1)+\overline{\sigma}_T^2$, $f_{i,j}=\overline{\sigma}_2^2-\overline{\sigma}_3^2$, $i\neq j$. If $\rho=1$ and $T=3$, $((\rho-r)P+I)'F^{-1}((\rho-r)P+I)=\Phi^{-1}$ and $\det(F)=\det(\Phi)$ hold for any value of r such that $-1<r\leq 1$, $f_{1,1}=\overline{\sigma}_2^2$, $f_{2,2}=\overline{\sigma}_2^2(1+\overline{\sigma}_3^2/\overline{\sigma}_2^2+r^2-2r)$ and $f_{1,2}=\overline{\sigma}_2^2(1-r)$.

It follows that $\Pr(\widetilde{\Delta} y_i : g(\widetilde{\Delta} y_i|r,F)\neq g(\widetilde{\Delta} y_i|\rho,\Phi))=1$ if $r\neq\rho$ and $f\neq\varphi$ unless $\rho=1$ and either $T<4$ or $\overline{\sigma}_s^2/\overline{\sigma}_{s-1}^2=\overline{\sigma}_3^2/\overline{\sigma}_2^2\neq 1$ for all $s\in\{3,...,T-1\}$. To establish unique identification in the last case, we can use similar arguments as at the end of the proof of theorem 1. We conclude that ρ and φ are uniquely identified if and only if $T\geq 3$ when $|\rho|<1$, or $T\geq 4$ when $\rho=1$.

$E\left(\sup_{\theta\in\Theta}\left|\ln g(\widetilde{\Delta} y_i|r,f)\right|\right)<\infty$ by standard arguments where $\theta=(r\ f')'$ and Θ is the parameter space. Finally $\bar{l}_{FE}(r,f)$ is continuous in r and f. We conclude that the FEQMLE for ρ is consistent if and only if $T\geq 3$ when $|\rho|<1$, or $T\geq 4$ when $\rho=1$.

Consistency of the REQMLE for ρ in the conditional panel AR(1) model can be proved along similar lines. However, instead of (35), one should use $y_i - ry_{i,-1}-p(1-r)y_{i,1}\iota = ((\rho-r)P+I)u_i+(\rho-r)P\iota(1-\rho)\pi y_{i,1}+(\pi(1-\rho)-p(1-r))y_{i,1}\iota$. □

References

[1] Ahn, S.C., Y.H. Lee and P. Schmidt, 2001, GMM estimation of linear panel data models with time-varying individual effects, Journal of econometrics 102, 219-255.

[2] Ahn, S.C., Y.H. Lee and P. Schmidt, 2013, Panel data models with multiple time-varying individual effects, Journal of econometrics 174, 1-14.

[3] Ahn, S.C., and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, Journal of econometrics 68, 5-28.

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

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

[6] Anderson, T.W., Kunitomo, N. and T. Sawa, 1982, Evaluation of the distribution function of the limited information maximum likelihood estimator, Econometrica 50, 1009-1027.

[7] Arellano, M., 2016, Modelling optimal instrumental variables for dynamic panel data models, Research in economics 70, 238-261.

[8] 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 LVIII, 277-297.

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

[10] Bai, J., 2013b, Likelihood approach to dynamic panel models with interactive effects, unpublished manuscript, Columbia University.

[11] Blundell, R.W. and R. Smith, 1991, Initial conditions and efficient estimation in dynamic panel data models: an application to company investment behaviour, Annales d'économie et de la statistique 20/21, 109-123.

[12] Bun, M.J.G., M.A. Carree and A. Juodis, 2017, On maximum likelihood estimation of dynamic panel data models, forthcoming in Oxford Bulletin of Economics and Statistics.

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

[14] Hayakawa, K., M.H. Pesaran and L.V. Smith, 2014, Transformed Maximum Likelihood Estimation of Short Dynamic Panel Data Models with Interactive Effects, unpublished manuscript, Hiroshima University, USC, Trinity College, Cambridge and York University.

[15] Holtz-Eakin, D., W. Newey and H. Rosen, 1988, Estimating vector autoregressions with panel data, Econometrica 56, 1371-1395.

[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] Hsiao, C., Zhang, J., 2015, IV, GMM or likelihood approach to estimate dynamic panel models when either N or T or both are large, Journal of econometrics 187, 312-322.

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

[19] Kruiniger, H., 2013wp, Quasi ML estimation of the panel AR(1) model with arbitrary initial conditions, working paper, Durham University, England.

[20] Kruiniger, H., 2016, Uniform Quasi ML based inference for the panel AR(1) model with arbitrary initial conditions, mimeo.

[21] Meng, X.-L. and D.B. Rubin, 1993, Maximum likelihood estimation via the ECM algorithm: a general framework, Biometrika 80, 267-278.

[22] Moral-Benito, E., 2013, Likelihood-based estimation of dynamic panels with predetermined regressors, Journal of business & economic statistics 31, 451-472.

[23] Mundlak, Y., 1978, On the pooling of time series and cross section data, Econometrica 46, 69-85.

[24] Newey, W., and D. McFadden, 1994, Large sample estimation and hypothesis testing, Ch. 36 in Handbook of econometrics, Vol. 4, R.F. Engle and D. McFadden (eds.), North-Holland, Elsevier, Amsterdam.

[25] Windmeijer, F., 2005, A finite sample correction for the variance of linear efficient two-step GMM estimators, Journal of econometrics 126, 25-51.

Table 1: Estimators and t-tests for ρ and β; Design I; 2500 replications.

design	A	N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
estimator		bias	MSE	bias	MSE	bias	MSE	bias	MSE
AB	ρ	-0.034	0.909	-0.008	0.153	-0.039	0.344	-0.009	0.052
	β	-0.015	0.570	-0.003	0.106	-0.026	0.241	-0.005	0.041
ABAS	ρ	-0.022	0.888	-0.005	0.140	-0.035	0.363	-0.003	0.042
	β	-0.012	0.629	-0.003	0.106	-0.023	0.288	-0.002	0.040
OPAS	ρ	-0.027	0.851	-0.006	0.141	-0.037	0.345	-0.005	0.043
	β	-0.014	0.595	-0.003	0.106	-0.024	0.254	-0.003	0.039
INFELIML	ρ	-0.016	0.354	-0.004	0.080	-0.004	0.112	-0.000	0.026
	β	-0.009	0.299	-0.002	0.062	-0.002	0.099	-0.001	0.021
ABASLIML	ρ	-0.022	0.414	-0.006	0.093	-0.012	0.134	-0.001	0.029
	β	-0.016	0.386	-0.004	0.081	-0.012	0.137	-0.001	0.027
OPASLIML	ρ	-0.023	0.418	-0.006	0.093	-0.013	0.134	-0.002	0.029
	β	-0.017	0.385	-0.004	0.081	-0.013	0.135	-0.002	0.027
FIML	ρ	0.005	0.622	0.001	0.123	0.002	0.160	0.000	0.030
	β	0.002	0.431	0.000	0.087	0.002	0.139	-0.000	0.025
		s.e.	rej f.	s.e.	rej.f.	s.e.	rej.f.	s.e.	rej.f.
AB	ρ	0.089	0.081	0.039	0.060	0.044	0.145	0.021	0.068
	β	0.074	0.063	0.032	0.050	0.041	0.115	0.019	0.066
INFELIML	ρ	0.059	0.065	0.029	0.064	0.033	0.072	0.016	0.060
	β	0.054	0.055	0.025	0.056	0.031	0.060	0.014	0.055
FIML	ρ	0.080	0.073	0.037	0.053	0.039	0.065	0.017	0.054
	β	0.068	0.070	0.032	0.058	0.036	0.069	0.016	0.046

s.e.: standard error; rej.f.: rejection frequency.

Table 2: Estimators and t-tests for ρ and β; Design I; 2500 replications.

design	B	N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
estimator		bias	MSE	bias	MSE	bias	MSE	bias	MSE
AB	ρ	-0.070	1.979	-0.013	0.301	-0.060	0.629	-0.013	0.080
	β	-0.007	0.604	-0.001	0.102	-0.015	0.214	-0.003	0.041
ABAS	ρ	-0.044	1.601	-0.007	0.255	-0.054	0.630	-0.004	0.057
	β	-0.006	0.679	-0.001	0.104	-0.015	0.277	-0.001	0.043
OPAS	ρ	-0.054	1.618	-0.009	0.258	-0.057	0.613	-0.007	0.059
	β	-0.007	0.642	-0.001	0.104	-0.015	0.237	-0.002	0.042
INFELIML	ρ	-0.031	0.645	-0.008	0.138	-0.010	0.148	-0.001	0.034
	β	-0.005	0.329	-0.001	0.068	-0.000	0.108	-0.000	0.020
ABASLIML	ρ	-0.035	0.676	-0.008	0.142	-0.015	0.162	-0.001	0.035
	β	-0.010	0.442	-0.002	0.091	-0.011	0.147	-0.001	0.028
OPASLIML	ρ	-0.036	0.680	-0.009	0.143	-0.015	0.163	-0.001	0.035
	β	-0.011	0.437	-0.002	0.091	-0.011	0.143	-0.001	0.028
FIML	ρ	0.013	1.366	0.004	0.247	-0.002	0.211	0.001	0.038
	β	0.001	0.443	0.001	0.093	0.002	0.143	-0.000	0.026
		s.e.	rej f.	s.e.	rej.f.	s.e.	rej.f.	s.e.	rej.f.
AB	ρ	0.123	0.107	0.053	0.061	0.053	0.202	0.025	0.086
	β	0.075	0.066	0.032	0.046	0.043	0.075	0.020	0.055
INFELIML	ρ	0.075	0.092	0.038	0.089	0.037	0.072	0.019	0.069
	β	0.057	0.059	0.026	0.058	0.032	0.061	0.014	0.055
FIML	ρ	0.110	0.126	0.054	0.063	0.045	0.066	0.020	0.053
	β	0.068	0.066	0.036	0.059	0.037	0.060	0.016	0.052

s.e.: standard error; rej.f.: rejection frequency.

Table 3: Estimators and t-tests for ρ and β; Design II; 2500 replications.

design	A	N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
estimator		bias	MSE	bias	MSE	bias	MSE	bias	MSE
AB	ρ	-0.038	1.137	-0.008	0.184	-0.046	0.425	-0.011	0.063
	β	-0.023	0.694	-0.003	0.124	-0.031	0.285	-0.007	0.045
ABAS	ρ	-0.027	1.101	-0.005	0.170	-0.041	0.438	-0.004	0.051
	β	-0.018	0.735	-0.002	0.125	-0.028	0.337	-0.003	0.044
OPAS	ρ	-0.032	1.080	-0.006	0.170	-0.043	0.421	-0.006	0.052
	β	-0.020	0.711	-0.003	0.125	-0.030	0.300	-0.004	0.044
INFELIML	ρ	-0.024	0.409	-0.007	0.085	-0.006	0.109	-0.001	0.025
	β	-0.014	0.339	-0.004	0.061	-0.003	0.101	-0.001	0.020
ABASLIML	ρ	-0.033	0.500	-0.009	0.106	-0.016	0.145	-0.002	0.031
	β	-0.024	0.475	-0.006	0.086	-0.016	0.148	-0.002	0.028
OPASLIML	ρ	-0.034	0.506	-0.009	0.105	-0.016	0.145	-0.003	0.031
	β	-0.026	0.478	-0.006	0.085	-0.017	0.146	-0.003	0.028
FIML	ρ	0.003	0.737	0.001	0.144	0.003	0.180	0.001	0.033
	β	-0.001	0.523	0.004	0.092	0.005	0.154	0.000	0.027
		s.e.	rej f.	s.e.	rej.f.	s.e.	rej.f.	s.e.	rej.f.
AB	ρ	0.096	0.090	0.042	0.054	0.047	0.163	0.023	0.077
	β	0.079	0.072	0.035	0.061	0.043	0.115	0.020	0.062
INFELIML	ρ	0.061	0.068	0.029	0.058	0.033	0.060	0.016	0.059
	β	0.056	0.069	0.026	0.043	0.031	0.057	0.015	0.048
FIML	ρ	0.094	0.080	0.046	0.060	0.043	0.066	0.018	0.050
	β	0.079	0.089	0.040	0.051	0.040	0.065	0.017	0.045

s.e.: standard error; rej.f.: rejection frequency.

Table 4: Estimators and t-tests for ρ and β; Design II; 2500 replications.

design	B	N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
estimator		bias	MSE	bias	MSE	bias	MSE	bias	MSE
AB	ρ	-0.086	2.574	-0.016	0.396	-0.070	0.789	-0.015	0.092
	β	-0.016	0.642	-0.004	0.127	-0.020	0.245	-0.005	0.047
ABAS	ρ	-0.050	1.913	-0.008	0.319	-0.062	0.770	-0.006	0.063
	β	-0.012	0.719	-0.003	0.128	-0.018	0.314	-0.002	0.049
OPAS	ρ	-0.064	1.948	-0.011	0.323	-0.066	0.759	-0.008	0.067
	β	-0.014	0.667	-0.004	0.128	-0.019	0.268	-0.003	0.048
INFELIML	ρ	-0.037	0.691	-0.013	0.153	-0.011	0.164	-0.003	0.032
	β	-0.003	0.363	-0.002	0.071	-0.001	0.108	0.000	0.022
ABASLIML	ρ	-0.043	0.748	-0.015	0.167	-0.018	0.191	-0.004	0.035
	β	-0.012	0.483	-0.004	0.104	-0.014	0.166	-0.001	0.032
OPASLIML	ρ	-0.044	0.754	-0.015	0.168	-0.018	0.192	-0.004	0.036
	β	-0.013	0.476	-0.004	0.104	-0.015	0.162	-0.002	0.032
FIML	ρ	0.016	1.593	0.002	0.296	0.003	0.282	-0.001	0.042
	β	0.003	0.462	0.000	0.101	0.003	0.166	0.001	0.029
		s.e.	rej f.	s.e.	rej.f.	s.e.	rej.f.	s.e.	rej.f.
AB	ρ	0.137	0.116	0.060	0.066	0.056	0.232	0.027	0.084
	β	0.079	0.061	0.034	0.065	0.045	0.075	0.021	0.062
INFELIML	ρ	0.076	0.085	0.039	0.076	0.037	0.093	0.019	0.066
	β	0.058	0.063	0.026	0.054	0.033	0.061	0.015	0.056
FIML	ρ	0.120	0.124	0.055	0.065	0.050	0.086	0.021	0.048
	β	0.071	0.073	0.033	0.064	0.041	0.083	0.017	0.054

s.e.: standard error; rej.f.: rejection frequency.

Table 5: Estimators and t-tests for ρ and β; Design III; 2500 replications.

design	A	N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
estimator		bias	MSE	bias	MSE	bias	MSE	bias	MSE
AB	ρ	-0.055	1.527	-0.010	0.251	-0.051	0.498	-0.012	0.073
	β	-0.028	0.752	-0.005	0.127	-0.031	0.282	-0.007	0.048
ABAS	ρ	-0.027	1.170	-0.005	0.192	-0.045	0.498	-0.005	0.052
	β	-0.018	0.773	-0.003	0.124	-0.029	0.334	-0.004	0.047
OPAS	ρ	-0.037	1.212	-0.007	0.195	-0.048	0.486	-0.007	0.054
	β	-0.023	0.749	-0.004	0.124	-0.030	0.297	-0.005	0.047
INFELIML	ρ	-0.001	0.635	0.002	0.134	0.000	0.156	-0.000	0.029
	β	-0.003	0.396	-0.000	0.080	0.001	0.112	-0.000	0.022
ABASLIML	ρ	-0.012	0.657	-0.000	0.149	-0.011	0.173	-0.002	0.034
	β	-0.014	0.478	-0.002	0.102	-0.013	0.150	-0.002	0.029
OPASLIML	ρ	-0.013	0.662	-0.001	0.149	-0.011	0.172	-0.002	0.034
	β	-0.016	0.482	-0.003	0.102	-0.014	0.148	-0.003	0.029
FIML	ρ	0.003	0.806	0.002	0.152	0.003	0.190	0.001	0.038
	β	-0.001	0.520	0.000	0.102	0.002	0.159	-0.000	0.026
		s.e.	rej f.	s.e.	rej.f.	s.e.	rej.f.	s.e.	rej.f.
AB	ρ	0.112	0.079	0.049	0.059	0.050	0.178	0.024	0.084
	β	0.082	0.079	0.036	0.052	0.043	0.120	0.020	0.062
INFELIML	ρ	0.079	0.098	0.037	0.050	0.039	0.056	0.017	0.050
	β	0.063	0.062	0.029	0.046	0.034	0.043	0.015	0.046
FIML	ρ	0.088	0.098	0.040	0.053	0.042	0.068	0.019	0.048
	β	0.072	0.077	0.032	0.053	0.039	0.065	0.017	0.047

s.e.: standard error; rej.f.: rejection frequency.

Table 6: Estimators and t-tests for ρ and β; Design III; 2500 replications.

design	B	N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
estimator		bias	MSE	bias	MSE	bias	MSE	bias	MSE
AB	ρ	-0.111	3.579	-0.023	0.513	-0.077	0.944	-0.018	0.114
	β	-0.017	0.644	-0.004	0.122	-0.019	0.227	-0.005	0.045
ABAS	ρ	-0.053	2.109	-0.008	0.366	-0.070	0.904	-0.006	0.065
	β	-0.011	0.730	-0.002	0.124	-0.017	0.306	-0.003	0.047
OPAS	ρ	-0.075	2.360	-0.013	0.373	-0.073	0.902	-0.010	0.071
	β	-0.014	0.677	-0.003	0.123	-0.018	0.256	-0.004	0.046
INFELIML	ρ	-0.028	0.835	-0.002	0.227	-0.003	0.190	0.000	0.044
	β	-0.002	0.354	-0.000	0.068	0.001	0.115	0.001	0.023
ABASLIML	ρ	-0.034	0.863	-0.004	0.229	-0.009	0.199	-0.001	0.046
	β	-0.011	0.478	-0.002	0.093	-0.012	0.164	-0.001	0.033
OPASLIML	ρ	-0.034	0.871	-0.004	0.230	-0.009	0.199	-0.001	0.046
	β	-0.012	0.471	-0.003	0.093	-0.013	0.157	-0.002	0.033
FIML	ρ	0.017	1.971	0.005	0.347	0.001	0.248	0.001	0.046
	β	0.001	0.463	0.001	0.093	0.003	0.174	-0.001	0.030
		s.e.	rej f.	s.e.	rej.f.	s.e.	rej.f.	s.e.	rej.f.
AB	ρ	0.153	0.138	0.068	0.074	0.060	0.250	0.029	0.089
	β	0.078	0.066	0.034	0.057	0.045	0.079	0.021	0.056
INFELIML	ρ	0.091	0.114	0.049	0.149	0.044	0.081	0.021	0.053
	β	0.060	0.054	0.027	0.044	0.033	0.056	0.015	0.053
FIML	ρ	0.125	0.139	0.059	0.063	0.049	0.061	0.022	0.054
	β	0.071	0.061	0.032	0.044	0.040	0.069	0.017	0.053

s.e.: standard error; rej.f.: rejection frequency.