Manuscript submitted to *The Econometrics Journal*, pp. 1–24.

Identification without assuming mean stationarity: Quasi ML estimation of dynamic panel models with endogenous regressors

Hugo Kruiniger [1] †

† *Dept. of Economics, Durham University, 1 Mill Hill Lane, Durham, DH1 3LB, UK.*
E-mail: hugo.kruiniger@durham.ac.uk

Summary Linear GMM estimators for dynamic panel models with predetermined or endogenous regressors suffer from a weak instruments problem when the data are highly persistent. In this paper we propose new random and fixed effects Limited Information Quasi ML estimators (LIQMLEs) for such models. We also discuss LIQMLEs for models that contain time-varying individual effects. Unlike System GMM estimators, the LIQMLEs do not require mean stationarity conditions for consistency. Such conditions often do not hold for the models we consider. Our LIQMLEs are based on a two-step control function approach that includes the first stage model residuals for a predetermined or endogenous regressor in the outcome equation. The LIMLEs are more precise than non-linear GMM estimators that are based on the original outcome equation. The LIQMLEs also compare favourably to various alternative (Q)MLEs in terms of precision, robustness and/or ease of computation.

Keywords: *control function, endogeneity, Generalized Method of Moments (GMM), Limited Information, predetermined regressors, Quasi Maximum Likelihood (QML), time-varying individual effects, weak identification.*

1. INTRODUCTION

In this paper we propose Random Effects (RE) and Fixed Effects (FE) Limited Information Quasi ML estimators for versions of the following panel AR(1) model with one additional regressor that is not strictly exogenous: [2]

$$y_{i,t} = \rho y_{i,t-1} + \beta x_{i,t} + \mu_i + \varepsilon_{i,t}, \tag{1.1}$$

for $i = 1, ..., N$ and $t = 2, ..., T$. [3] Specifically, we distinguish between the case where the regressor $x_{i,t}$ is predetermined with respect to the idiosyncratic error $\varepsilon_{i,t}$, i.e., $E(\varepsilon_{i,t}|y_i^{t-1}, x_i^t, \mu_i) = 0$, $t = 2, ..., T$, where $y_i^{t-1} = (y_{i,1} \ ... \ y_{i,t-1})'$ and $x_i^t = (x_{i,1} \ ... \ x_{i,t})'$; and the case where $x_{i,t}$ is contemporaneously correlated with $\varepsilon_{i,t}$, i.e., endogenous. [4] In addition, we allow $x_{i,t}$ to be correlated with the individual effect μ_i. We also discuss Quasi MLEs for models that include time-varying individual effects (also known as interactive

[1] I thank Anurag Banerjee, Jeff Wooldridge and Hong Il Yoo and especially Tom Wansbeek, Petra Todd (the co-editor) and two referees for helpful suggestions.

[2] Extensions to models with multiple lags and additional regressors are straightforward.

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

[4] The former case includes the case where $x_{i,t}$ is not just predetermined w.r.t. $\varepsilon_{i,t}$ but also weakly exogenous w.r.t. ρ and β.

effects or a factor structure), e.g.:

$$y_{i,t} = \rho y_{i,t-1} + \beta x_{i,t} + \delta_t \mu_i + \varepsilon_{i,t}, \tag{1.2}$$

for $i = 1, ..., N$ and $t = 2, ..., T$. We assume that $x_{i,1}$, $i = 1, ..., N$ are observed. The models considered allow for arbitrary initial conditions and heteroskedasticity. The asymptotic properties of the estimators are derived assuming $N \to \infty$ with T fixed. The FE estimators only exploit differenced data variation and hence can rely on minimal assumptions for their consistency whereas the RE estimators exploit data in levels.

Following Anderson and Hsiao (1982) and Arellano and Bond (1991), many researchers estimate a transformed version of a dynamic linear panel model using an instrumental variables (IV) approach or, more generally, the Generalized Method of Moments (GMM). [5] GMM estimators are appealing because they are semiparametric, do not invoke distributional assumptions and are easy to compute. However, these estimators suffer from a weak instruments problem when the data are persistent, that is, when the autoregressive parameter is close or equal to unity, cf. Bond and Blundell (1998). To alleviate this problem, the latter and Arellano and Bover (1995) proposed the so-called System GMM estimator which exploits additional moment conditions that rely on a mean stationarity assumption. However, in many applications this assumption does not hold.

Chamberlain (1980), Anderson and Hsiao (1981, 1982) and Bhargava and Sargan (1983) introduced several Random Effects maximum likelihood estimators (MLEs) for homoskedastic dynamic panel models with strictly exogenous regressors. Hsiao et al. (2002; henceforth HPT) proposed so-called Transformed MLEs for such models. They can be regarded as Fixed Effects MLEs. Alvarez and Arellano (2004) stressed that if there is time-series heteroskedasticity, then the MLEs need to allow for it to be consistent, while Kruiniger (2013) showed that MLEs for the panel AR(1) model that allows for heteroskedasticity over time remain consistent under arbitrary heteroskedasticity and non-normality, and, importantly, also when the autoregressive parameter is close or equal to unity. These MLEs do not require mean stationarity.

If the data distribution is correctly specified, then these MLEs have better finite sample properties than their GMM counterparts that exploit overidentifying moment conditions, including non-linear GMM estimators in the spirit of Ahn and Schmidt (1995), especially when the instruments are weak and/or many, cf. Anderson et al. (1982), Alvarez and Arellano (2003), Kruiniger (2013) and Hsiao and Zhang (2015). Of course, when the data distribution is incorrectly specified, then under first-order asymptotics the Quasi MLEs will be less precise than some of their optimal GMM counterparts. [6]

Motivated by these advantages of QML over GMM estimators, we first propose in section 2 RE LIQMLEs for dynamic panel models with predetermined regressors that can be correlated with the individual effect(s). Moral-Benito (2013) discusses what he calls subsystem LIMLEs (ssLIMLEs) for such models, which estimate them jointly with a set of reduced form equations for the predetermined regressors and the initial observations. [7] Bai (2013b) discusses Full Information (FI) QMLEs for such models; he considers LIQ-

[5] Holtz-Eakin et al. (1988) and Ahn et al. (2001, 2013) proposed GMM estimators for models with time-varying individual effects.

[6] Note, however, that the first-order asymptotic standard errors of such GMM estimators ignore the variation due to the presence of estimated parameters in the weight matrix and can be severely downward biased in finite samples, cf. Windmeijer (2005).

[7] However, Moral-Benito (2013) does not allow for interactive effects in the outcome equation.

MLEs for models with predetermined regressors that are weakly exogenous w.r.t. ρ and β. Our LIQMLEs are akin to a two-step control function approach, where residuals from first stage regressions for endogenous explanatory variables are included in the second stage model, cf. Wooldridge (2010, chap. 6). Unlike Moral-Benito (2013) and Bai (2013b), we also discuss FE QMLEs for the models considered in this paper. The FE LIQMLE is sometimes more efficient than the RE LIQMLE, e.g. in the case of an equation of a panel VAR model, when the individual time series are mean stationary and persistent and the variances of the initial conditions are small as compared to their values in case of a covariance stationary VAR process.

Our RE LIQMLE is more robust than Bai's RE FIQMLE: it remains consistent when the model of a predetermined regressor that is weakly exogenous w.r.t. ρ and β is misspecified, whereas the FIQMLE can become inconsistent in this case. Furthermore, we will argue in this paper that the global maxima of the likelihood functions associated with our LIQMLEs are more likely to be found than the global maxima of the likelihood functions associated with the FIQMLEs when the number of predetermined regressors (K) and the number of factors in the system of equations (R) are not small. Nevertheless, like the LIQMLEs, the FIQMLEs can be computed for long panels, i.e., panels with large T, whereas computation of the ssLIMLE requires that $(T-1) < N/K$, which suggests that the ssLIMLE is not suitable for panels with relatively large T.[8] Furthermore, as the ssLIMLE is partly based on reduced form equations for the predetermined regressors, consistency of the ssLIMLE requires that the lag length of these equations is not chosen to be too low. Finally, the Transformed MLE that HPT (2002) proposed for models with weakly exogenous regressors and time-invariant individual effects is also valid when the regressors are predetermined but not weakly exogenous. It can also be generalized to allow for heteroskedasticity. We will argue that it is less efficient but more robust than alternative MLEs, including ours, as no correctly specified model for $x_{i,t}$ is required.

In section 2 we also propose LIQMLEs for dynamic panel models with endogenous regressors.[9] [10] Because they include the first stage model residuals for an endogenous or a predetermined regressor in the outcome equation, our LIMLEs are more precise under normally distributed data than non-linear GMM estimators in the spirit of Ahn-Schmidt (1995). Table 1 below compares some of the aforementioned estimators for models with constant individual effects with our RE LIMLE in terms of their applicability and ease of use. When mean stationarity is likely to hold or the data are not close to normally distributed, a System GMM estimator or an Ahn-Schmidt type GMM estimator may be preferable to the RE LIQMLE.

In section 3 we examine the finite sample properties of several RE LIQML and GMM estimators, the ssLIMLE, the RE FIQMLE and some related Wald tests for fixed T dynamic panel models with predetermined regressors and time-invariant individual effects in a Monte Carlo study. We find that the RE LIMLEs for ρ (and β) usually have better finite sample properties than the ssLIMLE for ρ and β and the RE FIMLE for ρ. Using simulation experiments Moral-Benito (2013) has also compared the finite sample prop-

[8] Williams et al. (2017) note that their algorithm for computing the ssLIMLE works best when $T < 10$.

[9] The RE FIQMLE discussed in Bai (2013b) for models with predetermined regressors that are not weakly exogenous w.r.t. ρ and β can also be used for models with endogenous regressors.

[10] One can construct consistent LIQMLEs for static panel models with predetermined or endogenous regressors analogously to the LIQMLEs discussed in this paper.

erties of various ML and GMM estimators for such models, while Juodis and Sarafidis (2018) have compared the finite sample properties of two RE LIQMLEs of Bai (2013b) and various IV and GMM estimators for similar models with strictly or weakly exogenous regressors and interactive effects. Section 4 applies various estimators to a panel VAR model for employment and wage, and section 5 concludes.

Table 1. Comparison of various estimators for dynamic panel models.

	AB GMM	SYS GMM	HPT MLE	ssLIMLE	RE LIMLE we	RE FIMLE	RE LIMLE
Data are persistent & mean stationary		✓	✓	✓	✓	✓	✓
Data are not persistent & not mean stationary	✓		✓	✓	✓	✓	✓
Covariate(s) predetermined but not weakly exogenous	✓	✓	✓	✓		✓	✓
Model includes endogenous covariate(s)	✓	✓			✓	✓	✓
Large N and T	✓	✓	✓		✓	✓	✓
FE version available	✓		✓			✓	✓
Easy to compute	✓	✓	✓		✓		✓
Robust	✓		✓	✓			✓

✓: the estimator is (more) "suitable" in this case/regard (than (an) alternative estimator(s)); HPT MLE: version for predetermined regressors; LIMLE we: assumes weak-exogeneity.

2. THE ASSUMPTIONS AND THE ESTIMATORS

Throughout the paper we rely on the following **Basic Assumptions**: (i) T is fixed (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.

Because we assume that T is fixed, we do not need to restrict the parameter space for ρ and can allow $\rho > 1$. We could allow T to grow large if ρ is restricted to lie in $(-1, 1]$. When $\rho \geq 1$, the individual effect μ_i may not be part of the DGP, but we do not impose such a restriction on the models. Furthermore, 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 MLEs that we propose use the same variance parameters for all individuals, cf. Kruiniger (2013). The errors can also be conditionally heteroskedastic over time.

In this section we will first discuss QMLEs that are based on a single augmented equation, i.e., LIQMLEs for dynamic panel models with predetermined regressors and subsequently we will discuss FI- and LIQMLEs for models with endogenous regressors. The consistency proofs for the LIQMLEs 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 Online Appendix S.1 below. [11] [12] When $\rho = 1$, consistency of the QMLEs requires that $T \geq 4$ rather than $T \geq 3$, cf. Kruiniger (2013).

2.1. Limited information QML estimators for dynamic panel models with predetermined regressors

We first consider single equation based LIQML estimators for ρ and β in (1.1) or (1.2) when the $x_{i,t}$ are 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:

$$x_{i,t} = \alpha_x x_{i,t-1} + \beta_x y_{i,t-1} + \gamma_x \mu_i + \lambda_i + \xi_{i,t}, \tag{2.3}$$

where α_x, 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. 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 (1.1) and (1.2). A consistent RE LIQMLE 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 RE LIQML approach to estimating ρ and β in (1.1) when both $\beta_x \neq 0$ and $\gamma_x \neq 0$ in (2.3) but the $\xi_{i,t}$ are homoskedastic over time. Following the logic of the RE FIQML approach (cf. Bai, 2013b, and section 2.2 below) we first consider replacing μ_i in (1.1) by its projection on 1, $y_{i,1}$ and $x_{i,1}$, i.e., by

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

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

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

where $y_i = (y_{i,2} \, ... \, y_{i,T})'$, $y_{i,-1} = (y_{i,1} \, ... \, y_{i,T-1})'$, $x_i = (x_{i,2} ... x_{i,T})'$, $\iota = (1 \; 1 \; ... \; 1)'$, $\widetilde{u}_{y,i} = \widetilde{v}_{y,i}\iota + \varepsilon_i$, $\varepsilon_i = (\varepsilon_{i,2} \, ... \, \varepsilon_{i,T})'$ 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 $E(\mu_i \widetilde{v}_{y,i}) \neq 0$.

[11] These proofs assume $-1 < \rho \leq 1$ and that μ_i drops out of the model when $\rho = 1$. Fixed T, large N consistency of the QMLEs can still easily be shown when these restrictions are relaxed.

[12] Consistency of the QMLEs can be proved by using Theorem 2.1 in Newey and McFadden (1994, henceforth 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 would 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 Online Appendix S.1 below. The other conditions of Theorem 2.5 of NMcF, including the dominance condition, are easily verified, again see Online Appendix S.1.

However, one can obtain a consistent RE LIQML estimator for ρ and β in (1.1) by replacing μ_i in (1.1) by

$$\mu_y + \pi_y y_{i,1} + \varphi_y x_{i,1} + \psi_y \iota'(x_i - \widehat{\alpha}_x x_{i,-1} - \widehat{\beta}_x y_{i,-1}) + v_{y,i}, \tag{2.4}$$

where $\widehat{\alpha}_x$ and $\widehat{\beta}_x$ are preliminary consistent estimators of α_x and β_x. To see this, let $u_{y,i,t} = v_{y,i} + \varepsilon_{i,t}$ 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 $\gamma_x\mu_i + \lambda_i = \underline{\mu}_x + \underline{\pi}_x y_{i,1} + \underline{\varphi}_x x_{i,1} + \widetilde{v}_{x,i}$ 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 $\widetilde{\Phi}_{yx}\widetilde{\Phi}_{xx}^{-1} \propto \iota\iota'$, [13] one can interpret the augmented version of (1.1) in which μ_i is replaced by (2.4) 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 RE LIQMLE based on (1.1) with μ_i replaced by (2.4) 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 $\prod_{t=2}^{T} f(y_{i,t}|y_{i,t-1}, x_{i,t}, y_{i,1}, x_{i,1}, \widetilde{u}_{x,i})$. [14] [15] The consistency proof is similar to that given in Online Appendix S.1 for the RE QMLE 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 RE QMLE.

Our RE LIQMLE can be viewed as a two-step control function approach to estimation, where residuals from a first stage are inserted in the second stage estimation problem, cf. Wooldridge (2010, chapter 6). It is important to have $\widehat{\alpha}_x$ and $\widehat{\beta}_x$ in (2.4) rather than the unknown parameters α_x and β_x, because estimating ρ and β jointly with α_x and β_x by a single equation LIQML approach would result in an inconsistent estimator.

One can apply GMM to (2.3) to obtain the estimators $\widehat{\alpha}_x$ and $\widehat{\beta}_x$. Such GMM estimators can exploit the Arellano-Bond (1991) type moment conditions $E(x_{i,s}(\Delta x_{i,t} - \alpha_x\Delta x_{i,t-1} - \beta_x\Delta y_{i,t-1})) = 0$, $E(y_{i,s}(\Delta x_{i,t} - \alpha_x\Delta x_{i,t-1} - \beta_x\Delta y_{i,t-1})) = 0$, $s = 1, ..., t-2$, $t = 3, ..., T$, 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

$$y_{i,t} = \rho y_{i,t-1} + \beta x_{i,t} + \mu_y + \pi_y y_{i,1} + \varphi_y x_{i,1} + \psi_y\iota'(x_i - \widehat{\alpha}_x x_{i,-1} - \widehat{\beta}_x y_{i,-1}) + u_{y,i,t}, \tag{2.5}$$

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

$$x_{i,t} = \alpha_x x_{i,t-1} + \beta_x y_{i,t-1} + \mu_x + \pi_x y_{i,1} + \varphi_x x_{i,1} + \psi_x\iota'(y_i - \widehat{\rho} y_{i,-1} - \widehat{\beta} x_{i,-1}) + u_{x,i,t}, \tag{2.6}$$

and estimate these equations simultaneously by using the QML method while treating $\widehat{\rho}$, $\widehat{\beta}$, $\widehat{\alpha}_x$ 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 FIQML approach. However, simultaneously estimating (2.5) and (2.6) may not be entirely straightforward due to non-linearities. To

[13] Under homoskedasticity of the $\xi_{i,t}$ over time, $\widetilde{\Phi}_{xx}^{-1} = \sigma_\xi^{-2}(I_{T-1} - \iota\iota'/\iota'\iota) + c_x\iota\iota'$ for some constants σ_ξ^2 and c_x and hence $\widetilde{\Phi}_{yx}\widetilde{\Phi}_{xx}^{-1}$ is proportional to $\iota\iota'$.

[14] Note that $u_{y,i}$ depends on $\widehat{\alpha}_x$ and $\widehat{\beta}_x$, whereas $\underline{u}_{y,i} = \text{plim}_{N\to\infty} u_{y,i}$ depends on α_x and β_x.

[15] Note that given $y_{i,1}$, $x_{i,1}$ and $\widetilde{u}_{x,i}$, the elements of x_i can be recovered by using (2.3) and y_i.

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 RE LIQMLE can be based on (1.1) with μ_i replaced by (2.9) below. A FE LIQMLE for ρ and β in (1.1) when $x_{i,t}$ obeys (2.3) 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

$$\widetilde{\Delta} y_{i,t} = \rho \widetilde{\Delta} y_{i,t-1} + \beta \widetilde{\Delta} x_{i,t} + \mu_y + \psi_y \iota'(\widetilde{\Delta} x_i - \widehat{\alpha}_x \widetilde{\Delta} x_{i,-1} - \widehat{\beta}_x \widetilde{\Delta} y_{i,-1}) + u_{y,i,t}, \tag{2.7}$$

where the operator $\widetilde{\Delta}$ creates deviations from an initial observation, e.g. $\widetilde{\Delta} x_{i,-1} = x_{i,-1} - x_{i,1}\iota$, and $u_{y,i,t} = v_{y,i} + \varepsilon_{i,t}$. The FE model (2.7) is derived from the RE model (2.5) by applying $\widetilde{\Delta}$ to the observables in (2.5) and redefining $v_{y,i}$. One could estimate (2.3) by a suitable Transformed MLE of Hsiao et al. (2002) to obtain FE estimates of α_x and β_x

Next we discuss limited information RE QML estimation of ρ and β in the more general model (1.2) when, instead of (2.3), $x_{i,t}$, for its part, obeys the more general equation

$$x_{i,t} = \alpha_x x_{i,t-1} + \beta_x y_{i,t-1} + \gamma_t \mu_i + \vartheta_t \lambda_i + \xi_{i,t}, \tag{2.8}$$

where α_x, β_x, γ_t and ϑ_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 RE LIQMLE for ρ and β in (1.2) when $\beta_x \neq 0$ but $\gamma_t = 0$ for all t is given in Bai (2013b). Regardless of whether the $\xi_{i,t}$ are homo- or heteroskedastic, when $\beta_x \neq 0$ and $\gamma_t \neq 0$ for some or all t, a consistent RE LIQMLE for ρ and β can be obtained by applying the ML method to (1.2) with μ_i replaced by

$$\mu_y + \pi_y y_{i,1} + \varphi_y x_{i,1} + \sum_{t=2}^{T} \psi_{y,t}(x_{i,t} - \widehat{\alpha}_x x_{i,t-1} - \widehat{\beta}_x y_{i,t-1}) + v_{y,i}, \tag{2.9}$$

where $\widehat{\alpha}_x$ and $\widehat{\beta}_x$ are preliminary consistent estimators of α_x and β_x such as e.g. GMM estimators due of Ahn, Lee and Schmidt (2013). The terms added to (1.2) ensure that $\text{plim}_{N\to\infty} N^{-1} \sum_i ((\gamma_s \mu_i + \vartheta_s \lambda_i + \xi_{i,s}) u_{y,i,t}) = 0$, $\text{plim}_{N\to\infty} N^{-1} \sum_i (y_{i,1} u_{y,i,t}) = 0$ and $\text{plim}_{N\to\infty} N^{-1} \sum_i (x_{i,1} u_{y,i,t}) = 0$ for all s,t, where $u_{y,i,t} = \delta_t v_{y,i} + \varepsilon_{i,t}$, and effectively consistency of the RE LIQMLE for the parameters in (1.2), cf. the discussion in p. 6.

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

$$y_{i,t} = \rho y_{i,t-1} + \sum_{k=1}^{K} \beta_k x_{k,i,t} + \sum_{r=1}^{R} \delta_{r,t} \mu_{r,i} + \varepsilon_{i,t}, \tag{2.10}$$

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

$$x_{k,i,t} = \alpha_{x,k} x_{k,i,t-1} + \beta_{x,k} y_{i,t-1} + \sum_{r=1}^{R} \gamma_{k,r,t} \mu_{r,i} + \vartheta_{k,t} \lambda_{k,i} + \xi_{k,i,t}, \tag{2.11}$$

where $\alpha_{x,k}$, $\beta_{x,k}$, $\gamma_{k,r,t}$ and $\vartheta_{k,t}$ 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. To achieve identification of (2.10) one could impose R^2 restrictions on $\delta_{r,t}$, $r = 1,\dots,R$ and $t = 2,\dots,T$. A well-known identification strategy imposes $\delta_{r,r+1} = 1$ and $\delta_{r,s+1} = 0$ for $r \neq s$ and $1 \leq r,s \leq R$, see e.g. Bai and Li (2012). A similar comment applies to (2.11). A consistent RE LIQMLE for (2.10) can be obtained by applying the ML method to it with $\mu_{r,i}$

replaced by

$$\mu_{y,r}+\pi_{y,r}y_{i,1}+\varphi_{y,r}x_{i,1}+\sum_{k=1}^{K}\sum_{t=2}^{T}\psi_{y,k,r,t}(x_{k,i,t}-\widehat{\alpha}_{x,k}x_{k,i,t-1}-\widehat{\beta}_{x,k}y_{i,t-1})+v_{y,r,i}, \quad (2.12)$$

where $\widehat{\alpha}_{x,k}$ and $\widehat{\beta}_{x,k}$ are preliminary consistent estimators of $\alpha_{x,k}$ and $\beta_{x,k}$, $k=1,2,...,K$. 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 (2.3), (2.8) or (2.11) first and then test whether $\beta_{x(,k)}=0$.

Remark 2: 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 (2.4), (i.e., from (2.5)), (2.9) or (2.12) without causing inconsistency of the RE QMLE, 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 (2.7) or one of its generalizations would cause inconsistency of the FE QMLE 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 3: Unless x is weakly exogenous, consistency of the LIMLE for ρ and β requires that not only the model for y but also the model for x is correct. One can use the Sargan test to test for misspecification of these models.

Remark 4: The Transformed MLE of Hsiao et al. (2002) for models with predetermined regressors is less efficient than our FE MLEs, e.g. the FE MLE based on (2.7). This can be seen as follows. The Transformed MLE is based on the following system of equations:

$$\begin{aligned}\Delta y_{i,2}&=\delta_0^*+\delta_1^*\Delta x_{i,2}+\xi_i^*+\varepsilon_{i,2}, \qquad (2.13)\\ \Delta y_{i,t}&=\rho\Delta y_{i,t-1}+\beta\Delta x_{i,t}+\Delta\varepsilon_{i,t}, \quad t=2,...,T,\end{aligned}$$

where $E(\xi_i^*)=0$ and $E(\xi_i^*\varepsilon_{i,t})=0$, $\quad t=2,...,T$, and Gaussianity of the error components. The system of equations in (2.7) can be rewritten as the system in (2.13) with $\delta_1^*\Delta x_{i,2}$ in the equation for $\Delta y_{i,2}$ replaced by $\beta\Delta x_{i,2}+\psi_y\iota'(\widetilde{\Delta}x_i-\widehat{\alpha}_x\widetilde{\Delta}x_{i,-1}-\widehat{\beta}_x\widetilde{\Delta}y_{i,-1})$. Thus in the case of our FE MLE, the equation for $\Delta y_{i,2}$ also provides information on β and hence, when T is fixed, our FE MLE for β and ρ is more precise than the Transformed MLE. On the other hand, consistency of the Transformed MLE does not depend on a correctly specified model for $\widetilde{\Delta}x$ and hence this estimator is more robust in this respect than our FE MLE. Both estimators are restricted versions of an estimator that is based on

$$\begin{aligned}\Delta y_{i,2}&=\delta_0^*+\delta_1^*\Delta x_{i,2}+\psi_y\iota'(\widetilde{\Delta}x_i-\widehat{\alpha}_x\widetilde{\Delta}x_{i,-1}-\widehat{\beta}_x\widetilde{\Delta}y_{i,-1})+\xi_i^*+\varepsilon_{i,2}, \qquad (2.14)\\ \Delta y_{i,t}&=\rho\Delta y_{i,t-1}+\beta\Delta x_{i,t}+\Delta\varepsilon_{i,t}, \quad t=2,...,T.\end{aligned}$$

If the hypothesis $\delta_1^*=\beta$ is rejected, one should not use the FE MLE based on (2.7). Similar comments apply to versions of these estimators that are valid under heteroskedasticy.

Remark 5: Our LIQMLEs are (much) easier to compute than the ssLIMLE of Moral-Benito (2013). Furthermore, although the latter uses reduced form models for the regressors, the lag length of x and y in these models is restricted by the availability of data.

Remark 6: the LIQMLEs discussed in this section are based on augmented models for y that include generated regressors, namely the residuals of the models for the regressors. To compute standard errors for the LI (Quasi) MLEs for ρ and β_k, $k = 1, ..., K$, one can make use of (a sandwich version of) the formula for the limiting variance of two-step MLEs that is given in equation (34) in Murphy and Topel (1985) with $R_4 = 0$. One can also use the bootstrap to compute standard errors for the LIQMLEs. The latter approach is particularly attractive when there are several predetermined regressors.

2.2. QML estimators for dynamic panel models with endogenous regressors

The preceding limited information RE and FE QML estimators will no longer be consistent for ρ and β in (1.1) or (1.2) 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 (2.3) or $\gamma_t = 0$, $t = 2, ..., T$ in (2.8). Note that in both cases, $x_{i,t}$ is also still affected by lags of $\varepsilon_{i,t}$ through $y_{i,t-1}$. In these cases, we can adopt a Full Information QML approach to estimation that is based on a VAR model, cf. Bai (2013b). Upon substituting the RHS of (2.8) for $x_{i,t}$ in (1.2) and letting $\xi_{i,t}$ absorb $\vartheta_t \lambda_i$, we obtain the following VAR model:

$$\begin{bmatrix} y_{i,t} \\ x_{i,t} \end{bmatrix} = \begin{bmatrix} \rho + \beta_x \beta & \alpha_x \beta \\ \beta_x & \alpha_x \end{bmatrix} \begin{bmatrix} y_{i,t-1} \\ x_{i,t-1} \end{bmatrix} + \begin{bmatrix} (\delta_t + \gamma_t \beta) \\ \gamma_t \end{bmatrix} \mu_i + \begin{bmatrix} \varepsilon_{i,t} + \beta \xi_{i,t} \\ \xi_{i,t} \end{bmatrix}. \quad (2.15)$$

This model can be written more succinctly as

$$z_{i,t} = A z_{i,t-1} + \zeta_t \mu_i + \omega_{i,t}, \quad (2.16)$$

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 RE FIQMLE for ρ and β in (2.16), 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

$$z_{i,t} = A z_{i,t-1} + \zeta_t(\mu + \pi y_{i,1} + \varphi x_{i,1}) + u_{i,t}, \quad (2.17)$$

where $u_{i,t} = \zeta_t v_i + \omega_{i,t}$.[16] To achieve identification we can impose $\sigma_v^2 = 1$. Note that not only the ζ_t but also β 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 β is only identified by the mean equations and the FIQMLE for ρ and β 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 it is more practical to apply the FIQMLE to the system that consists of (1.2) and (2.8) with μ_i replaced by $\mu + \pi y_{i,1} + \varphi x_{i,1} + v_i$ rather than to (2.17).

[16] 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 $\beta \underline{\tau}_3 y_{i,1}$ to the equation for $y_{i,3}$, the term $\underline{\tau}_3 y_{i,1}$ to the equation for $x_{i,3}$, the term $\beta(\underline{\tau}_2 y_{i,1} + \underline{\vartheta}_2 x_{i,1})$ to the equation for $y_{i,2}$ and the term $\underline{\tau}_2 y_{i,1} + \underline{\vartheta}_2 x_{i,1}$ to the equation for $x_{i,2}$.

The model in (2.16) can also be estimated by a FE FIQMLE. In this case we estimate the system

$$\widetilde{\Delta}z_{i,t} = A\widetilde{\Delta}z_{i,t-1} + \underline{\lambda} + \underline{\lambda}_i + \zeta_t\mu_i + \omega_{i,t}, \tag{2.18}$$

where $\widetilde{\Delta}z_{i,t} = z_{i,t} - z_{i,1}$, $\underline{\lambda} = (\lambda_1\ \lambda_2)'$ and $\underline{\lambda}_i = (\lambda_{1,i}\ \lambda_{2,i})'$. To identify the model we can impose $\sigma^2_\mu = 1$.

If $Cov(\xi_{i,t}, \varepsilon_{i,t}) = 0$ for some or all t, then it is again more practical to apply the FIQMLE to a transformed version of the original system, which consists of

$$\widetilde{\Delta}y_{i,t} = \rho\widetilde{\Delta}y_{i,t-1} + \beta\widetilde{\Delta}x_{i,t} + \lambda_3 + \lambda_{3,i} + \delta_t\mu_i + \varepsilon_{i,t} \text{ and} \tag{2.19}$$
$$\widetilde{\Delta}x_{i,t} = \alpha_x\widetilde{\Delta}x_{i,t-1} + \beta_x\widetilde{\Delta}y_{i,t-1} + \lambda_2 + \lambda_{2,i} + \gamma_t\mu_i + \xi_{i,t}.$$

Consistency of the FIQMLE for ρ and β (and α_x and β_x) in (2.17) (or (2.18)) can be shown similarly to the RE LIQMLE 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 (2.17) to achieve consistency rather than efficiency; estimating the equations in (2.17) separately by ML would result in an inconsistent estimator of the parameters unless $\gamma_t = 0$, $t = 2, ..., T$.

Bai (2013b) describes the ECM algorithm of Meng and Rubin (1993) for computing QMLEs based on likelihood functions very similar to (2.17). The algorithm can be applied directly to (2.17) when $Cov(\xi_{i,t}, \varepsilon_{i,t}) \neq 0$ for all t or to (1.2) and (2.8) with μ_i replaced by $\mu + \pi y_{i,1} + \varphi x_{i,1} + v_i$ when $Cov(\xi_{i,t}, \varepsilon_{i,t}) = 0$ for at least some t, see remark 10 below for further discussion.

As an alternative to the above estimators, we can generalize the two-step LIQMLEs of section 2.1 to control for endogeneity of a regressor, i.e., correlation between, say, $x_{k,i,t}$ or $\widetilde{\Delta}x_{k,i,t}$ and the idiosyncratic error term, by adding the composite residual of the equation for $x_{k,i,t}$ or $\widetilde{\Delta}x_{k,i,t}$ as a regressor to the equation for $y_{i,t}$ or $\widetilde{\Delta}y_{i,t}$ for each $t \in \{2, \ldots, T\}$.

Remark 7: The standard errors for the FIQMLEs can be computed by using a sandwich formula. The standard errors for the LIQMLEs can be computed by using equation (34) in Murphy and Topel (1985) with $R_4 \neq 0$ or the bootstrap, cf. remark 6.

Remark 8: The FIQML approach can, of course, also be used for models with predetermined regressors where $Cov(\xi_{k,i,t}, \varepsilon_{i,s}) = 0$ for all $k \in \{1, ..., K\}$ and $s, t \in \{2, ..., T\}$ and offers an alternative to the LIQMLEs that have been discussed in section 2.1. When the models are correctly specified and the data are i.i.d. across the individuals and Gaussian, both the FIMLEs and the LIMLEs for ρ and β are consistent and asymptotically efficient. However, the LIQMLE and FIQMLE have different finite sample 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 and the total number of factors are not small, see remark 10 below.

Remark 9: The RE LIQMLEs for ρ and β in models that contain predetermined regressors with $Cov(\xi_{k,i,t}, \varepsilon_{i,s}) = 0$ for all $s, t \in \{2, ..., T\}$ and possibly some endogenous regressors are more robust than the RE FIQMLEs for ρ and β because they remain consistent when the model for a regressor that is weakly exogenous w.r.t. ρ and β is misspecified, whereas the FIQMLEs will become inconsistent in that case if this regressor is not weakly exogenous w.r.t. the slope parameters in the model for another regressor of the model for y that is not weakly exogenous w.r.t. ρ and β. The reason for the robustness of the RE LIQMLE is that, unlike the RE FIQMLE, the LI method estimates all the equations in the system separately and thereby prevents misspecification of a model for a regressor that is weakly exogenous w.r.t. ρ and β from causing inconsistency of the estimator for these parameters through spillover effects.

Remark 10: The likelihood functions associated with the FI- and LIQMLEs for models with interactive effects are nonlinear in the parameters and have multiple maxima. In particular, note that the mean equations are bilinear in the parameters. Note also that the factors appear both in the mean equation(s) and in the covariance matrix of the composite errors and that consistent QML estimation of the slope parameters depends on consistent estimation of the covariance matrix of the composite errors. The ECM algorithm is only guaranteed to converge to a local maximum of the likelihood function. To ensure consistency of the QMLE that is computed by using the ECM algorithm, one could use consistent starting values. As part of a method to obtain the latter, one can use a consistent version of the GMM estimator of Ahn, Lee and Schmidt (2013) that relies on a generalization of the CCE approach introduced by Pesaran (2006) and grid search over the autoregressive parameter(s) to estimate the factors and the slope parameters, see Kruiniger (2019) for details. Next one can use OLS to estimate the projection parameters. The composite residuals and estimates of the factors can then be used to obtain starting values for the elements of the covariance matrix of the projection errors ('loadings') and the idiosyncratic variances. However, the higher the number of equations and the total number of factors in the system, the higher the number of local maxima of the likelihood function for the full system [17] and the higher the probability of not converging to its global maximum even if one uses consistent starting values. In such a situation using the LIQMLE is appealing. When computing the LIQMLE, the ECM algorithm is applied to a smaller estimation problem, one avoids having to estimate the covariance matrix of the 'loadings' of all the factors in the system and one also avoids estimating the projection parameters in the equations for the regressors. Note also that the GMM estimates that are used by the ECM algorithm as starting values to compute the FIQMLE are also used by the the LIQMLEs either as starting values or to construct the controls for endogeneity. This suggests that the cost of computing the LIQMLEs is lower than that of the FIQMLE, especially when the system is not small.

Remark 11: 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 12: To test whether $y_{i,t}$ is affected by $x_{i,t-1}$ rather than by $x_{i,t}$ one can follow 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. $y_{i,t} = \lambda(\widehat{\rho} y_{i,t-1} + \widehat{\beta} x_{i,t}) + (1-\lambda)(\widehat{\rho}_{(t-1)} y_{i,t-1} + \widehat{\beta}_{(t-1)} x_{i,t-1}) + \delta_t(\mu + \pi y_{i,1} + \varphi x_{i,1} + \psi_y \iota'(x_i - \widehat{\alpha}_x x_{i,-1} - \widehat{\beta}_x y_{i,-1})) + u_{i,t}$ if one assumes that $Cov(\xi_{i,t}, \varepsilon_{i,s}) = 0$ for all $s, t \in \{2, ..., T\}$, where $\widehat{\rho}_{(t-1)}$ and $\widehat{\beta}_{(t-1)}$ are estimators for ρ and β in the model in which $y_{i,t}$ is affected by $x_{i,t-1}$ rather than by $x_{i,t}$ and $u_{i,t} = \delta_t v_i + \eta_{i,t}$ 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 13: Once the correct lag structures of x and y in the equation for $y_{i,t}$ have been determined, one can proceed to test for endogeneity of the regressor(s).

[17] This is suggested by a version of Bézout's theorem.

3. THE FINITE SAMPLE PERFORMANCE OF THE ESTIMATORS

In this section we compare through Monte Carlo simulations the finite sample properties of several RE LI- and FIMLEs and GMM estimators for the slope coefficients in the first equation of 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 while 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

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

We conducted the simulation experiments for $(T,N) = (5,100)$, $(10,100)$, $(5,500)$ and $(10,500)$ and five combinations of slope coefficients: A) $\rho = \alpha_x = 0.2$ and $\beta = \beta_x = 0.6$; B) $\rho = \alpha_x = 0.6$ and $\beta = \beta_x = 0.2$; C) $\rho = \alpha_x = 0.75$ and $\beta = \beta_x = 0.2$; D) $\rho = \alpha_x = 0.9$ and $\beta = \beta_x = 0.1$; and E) $\rho = 0.75$, $\alpha_x = 0.4$ and $\beta = \beta_x = 0.2$. In all simulation experiments $(\varepsilon_i'\ \xi_i')' \sim i.i.d.\ \mathrm{N}(0, diag(\sigma_\varepsilon^2, \sigma_\xi^2) \otimes I_{T-1})$ with $\xi_i = (\xi_{i,2}\ ...\ \xi_{i,T})'$ and $\sigma_\varepsilon^2 = \sigma_\xi^2 = 1$, and the vectors of the error components $(\mu_{y,i}\ \mu_{x,i})'$ and $(\varepsilon_j'\ \xi_j')'$ are mutually independent for all $i,j \in \{1,...,N\}$. Given the values of N and T and ρ, α_x, β and β_x we considered four designs of the experiments:

I) $\{(y_{i,t}, x_{i,t})\}$ is covariance stationary and $\mu_{y,i} = \mu_{x,i} \sim i.i.d.\ \mathrm{N}(0,\sigma_\mu^2)$ with $\sigma_\mu^2 = 1$. This means that there exists a matrix R such that $(y_{i,1}-\mu_{y,i}\ x_{i,1}-\mu_{x,i})' = R(\underline{\varepsilon}_{i,1}\ \underline{\xi}_{i,1})'$, where $(\underline{\varepsilon}_{i,1}\ \underline{\xi}_{i,1})' \sim i.i.d.\ \mathrm{N}(0, diag(\sigma_\varepsilon^2, \sigma_\xi^2))$.

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

III) $y_{i,1} = \mu_{y,i} + \underline{\varepsilon}_{i,1}$, $x_{i,1} = \mu_{x,i} + \underline{\xi}_{i,1}$, $\mu_{y,i} = 0.8 * \mu_{z,i} + 0.6 * \lambda_{y,i}$ 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 $(\underline{\varepsilon}_{i,1}\ \underline{\xi}_{i,1})' \sim i.i.d.\,\mathrm{N}(0, \varsigma * diag(\sigma_\varepsilon^2, \sigma_\xi^2))$, where $\varsigma = 1$.

IV) $(y_{i,1}-\mu_{y,i}\ x_{i,1}-\mu_{x,i})' = R(\mu_{y,i} + \underline{\varepsilon}_{i,1}\ \mu_{x,i} + \underline{\xi}_{i,1})'/\sqrt{2}$, where R, $\mu_{y,i}$, $\underline{\varepsilon}_{i,1}$, $\mu_{x,i}$ and $\underline{\xi}_{i,1}$ are the same as under design I).

Note that in designs II and III $\{(y_{i,t}\ x_{i,t})\}$ is mean stationary but not covariance stationary. In design IV $\{(y_{i,t}\ x_{i,t})\}$ is no longer mean stationary.

We also considered variations of designs II and III, viz. II$'$ and III$'$ where $\varsigma = 1/9$, a variation of design IV, viz. IV$'$ where $(y_{i,1}-\mu_{y,i}\ x_{i,1}-\mu_{x,i})' = R(\mu_{y,i}+\underline{\varepsilon}_{i,1}\ \mu_{x,i}+\underline{\xi}_{i,1})' * \sqrt{2}$ and $\mu_{y,i} = \mu_{x,i} \sim i.i.d.\ \mathrm{N}(0,5)$, and another variation of design IV, viz. IV$''$ where $\mu_{y,i} \sim i.i.d.\ \mathrm{N}(0,9)$ and $\mu_{x,i} = 0.5\mu_{y,i}$.

We considered the following estimators for ρ and β: a single equation two-step optimal Arellano-Bond (AB) type GMM estimator based on the linear moment conditions given in section 2.1; two versions of a single equation multi-step optimal Ahn-Schmidt (AS) type GMM estimator as described in section 2.1, 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; two estimators, OPASH and ABASH, that are very similar to OPAS and ABAS but also exploit moment conditions implied by homoskedasticity;

four versions of the RE LIMLE, i.e., an infeasible version that replaces $\widehat{\alpha}_x$ and $\widehat{\beta}_x$ by the true values of α_x and β_x, INFLIML, a version that replaces $\widehat{\alpha}_x$ and $\widehat{\beta}_x$ by AB estimates, ABLIML, and two versions that replace $\widehat{\alpha}_x$ and $\widehat{\beta}_x$ by one of the aforementioned AS GMM estimates for α_x and β_x, OPASLIML and ABASLIML; a general version of the ssLIMLE and, for $T = 5$ only, a version with homoskedasticity imposed, viz. ssLIMLh; a FE (a Transformed) MLE of HPT (2002) for models with predetermined regressors; a FE LIMLE that replaces $\widehat{\alpha}_x$ and $\widehat{\beta}_x$ by HPT ML estimates; and finally the RE FIMLE. A list of these estimators is given in table 2 together with a brief description of each estimator. The aforementioned AS type GMM estimators use AB residuals in the third step to estimate the optimal weight matrix. Note that these AS type GMM estimators are less precise than the MLEs because they ignore some moment conditions that are (partly) based on the model for the regressor.

In the experiments we allowed for time effects by subtracting cross-sectional averages from the data. 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 $\sigma^2_\varepsilon > 0$ and $Var(\widetilde{v}_{y,i}) > 0$ in the LI case and $\sigma^2_\varepsilon > 0$, $\sigma^2_\xi > 0$, $\sigma^2_\xi + (T-1)Var(\widetilde{v}_{x,i}) > 0$ and $(\sigma^2_\xi + (T-1)Var(\widetilde{v}_{x,i}))(\sigma^2_\varepsilon + (T-1)Var(\widetilde{v}_{y,i})) - (T-1)^2(Cov(\widetilde{v}_{x,i}, \widetilde{v}_{y,i}))^2 > 0$ in the FI case. The restrictions in the LI case are stronger than those in the FI case and could be relaxed to $\sigma^2_\varepsilon > 0$ and $\sigma^2_\varepsilon + (T-1)Var(\widetilde{v}_{y,i}) > 0$. 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. [18]

As the models used in the experiments only contain time-invariant individual effects, the mean equations are linear in the parameters and there are no restrictions between the parameters in the mean equation and the parameters in the covariance matrix. In this case computation of the MLEs is easy as compared to the case with interactive effects, cf. remark 10. For instance, using the zig-zag algorithm in Oberhofer and Kmenta (1974) with consistent starting values for the slope parameters would lead to consistent LI- and FIMLEs. We used the Constrained Maximum Likelihood module of the GAUSS software package to compute these MLEs, and the STATA programme xtdpdml of Williams et al. (2017) to compute the ssLIMLEs.

For the estimators we calculated the bias and the Mean Squared Error (MSE) and in some cases also the average standard error (s.e.). The s.e. of the AB estimator is based on Windmeijer's (2005) formula. The s.e. of the ABLIMLE is based on Murphy and Topel's (1985) formula. We also computed the empirical size, i.e., rejection frequency (rej.f.) of Wald tests based on the AB, the INFLIML, the ABLIML and the FIML

[18] 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.

estimators, respectively. All tests had a nominal size of 5%. Note that in practice it would be preferable to use robust (i.e. uniform) inference procedures, cf. Kruiniger (2016).

The simulation results for the various experimental designs are reported in seventeen tables in total: six tables corresponding to slope parameter combinations C and E are included in the main text, while a further eleven tables containing the results for the slope parameter combinations A, B and D and some additional results for experiments II-C and III-C are included in the Online Appendix. The MSE of the estimators has been multiplied by 100. For the ssLIMLE the results are based on cases where convergence has been achieved. Inspection of the results leads to the following conclusions:

1 The MSE of the OPASH (ABASH) GMM estimator is similar to or higher than that of the OPAS (ABAS) GMM estimator (the results for ABASH are not reported).

2 Our RE LIMLEs have much better finite sample properties, i.e., smaller bias and MSE than the AB, ABAS and OPAS GMM estimators. The properties of the RE LIMLEs are hardly affected by the choice among the preliminary GMM estimators that are used to estimate α_x and β_x.

3 Our RE LIMLEs are often (much) more precise, i.e., have a (much) smaller MSE than the FIMLEs, with the exception of the RE LIMLEs for β in designs C and D.

4 The SYS estimators have a higher or similar MSE than our RE LIMLEs in design I-A when $T = 5$, in some cases in design I-B, in design III-A when $T = 10$, and, unsurprisingly perhaps, in designs IV-A, IV-B, IV$'$-C and IV$''$-E, for which the SYS estimator is inconsistent. In all other designs, including IV-C, the SYS estimators have a lower MSE than our RE LIMLEs. Although the SYS estimator is inconsistent in designs IV-C and IV$'$-C, it is only mildly biased in these designs where the data are fairly persistent.

Table 2. Description of estimators.

Abbreviation	Description
AB	2-step optimal Arellano-Bond GMM estimator
ABAS	4-step optimal linearized Ahn-Schmidt GMM estimator using AB estimates in optimal weight matrix at step 3 and in differenced part of nonlinear moment conditions
OPAS	3-step optimal nonlinear Ahn-Schmidt GMM estimator using AB estimates in optimal weight matrix at step 3
OPASH	similar to OPAS estimator but exploiting extra moment conditions that are implied by assumption of homoskedasticity over time
INFLIML	RE LIMLE using true values to compute the residuals (the controls)
ABLIML	RE LIMLE using AB estimates to compute the residuals (the controls)
ABASLIML	RE LIMLE using ABAS estimates to compute the residuals (the controls)
OPASLIML	RE LIMLE using OPAS estimates to compute the residuals (the controls)
ssLIML	ssLIMLE of Moral-Benito (without assuming homoskedasticity)
ssLIMLh	ssLIMLE of Moral-Benito that imposes homoskedasticity
HPT	FE MLE of HPT (2002) for model with predetermined regressors
HPTFLIML	FE LIMLE using HPT estimates to compute the residuals (the controls)
FIML	a RE FIMLE of Bai (2013b) that imposes homoskedasticity

5 The ssLIMLh estimators tend to have a slightly higher MSE than the ssLIMLEs except in design III-B.
6 The ssLIMLEs usually have a larger MSE than our RE LIMLEs except in designs II-C and II-D when $T = 10$, and in the case of the ssLIMLE for β, in design III-C when $T = 10$.
7 The STATA algorithm used to compute the ssLIMLEs often did not converge.
8 Our HPT MLE based FE LIMLEs have a higher MSE than our RE LIMLEs except in the following designs: the A versions of designs I, II and III when $T = 10$ and $N = 100$, and in the case of the FE LIMLE for β, design IV-B, the C versions

Table 3. Estimators and t-tests for ρ and β; Design I-C; 2500 replications.

		N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
estimator	param.	bias	MSE	bias	MSE	bias	MSE	bias	MSE
AB	ρ	-.094	3.134	-.017	.453	-.085	1.030	-.019	.116
	β	-.018	.585	-.005	.108	-.022	.229	-.005	.046
ABAS	ρ	-.072	2.506	-.010	.406	-.079	.997	-.010	.089
	β	-.020	.693	-.005	.114	-.020	.277	-.003	.049
OPAS	ρ	-.079	2.499	-.013	.403	-.082	.997	-.013	.092
	β	-.020	.634	-.005	.112	-.021	.244	-.004	.048
SYS	ρ	-.009	.626	-.001	.131	-.005	.105	-.002	.032
	β	.002	.457	.001	.088	.003	.095	.001	.029
INFLIML	ρ	-.050	.808	-.018	.171	-.021	.145	-.007	.027
	β	-.004	.327	-.001	.061	-.001	.084	-.001	.016
ABLIML	ρ	-.058	.900	-.021	.190	-.032	.215	-.010	.040
	β	-.018	.468	-.005	.094	-.024	.177	-.007	.036
ABASLIML	ρ	-.059	.899	-.021	.188	-.031	.212	-.009	.037
	β	-.018	.481	-.004	.095	-.022	.180	-.005	.035
OPASLIML	ρ	-.059	.900	-.021	.189	-.031	.213	-.010	.038
	β	-.018	.470	-.005	.094	-.023	.176	-.006	.035
ssLIML	ρ	-.045	1.564	-.028	.319	-.014	.293	-.004	.045
	β	-.001	.548	-.003	.115	-.001	.191	-.001	.034
HPTFLIML	ρ	-.025	1.453	-.011	.514	-.028	.265	-.023	.095
	β	-.041	.490	-.034	.178	-.034	.207	-.035	.134
HPT	ρ	.075	4.296	.002	1.257	-.028	.578	-.031	.138
	β	-.068	.734	-.063	.421	-.059	.485	-.058	.344
FIML	ρ	.006	1.586	.003	.359	-.003	.252	.000	.049
	β	-.003	.459	-.001	.088	-.001	.130	-.001	.029
		s.e.	rej.f.	s.e.	rej.f.	s.e.	rej.f.	s.e.	rej.f.
AB	ρ	.140	.122	.060	.071	.056	.324	.027	.115
	β	.073	.064	.032	.048	.041	.090	.019	.072
INFLIML	ρ	.052	.101	.020	.062	.031	.096	.019	.070
	β	.055	.062	.024	.062	.029	.052	.013	.060
ABLIML	ρ	NA	.058	.059	.031	.049	.049	.023	.027
	β	NA	.012	.032	.017	.037	.055	.018	.041
FIML	ρ	.098	.152	.049	.085	.044	.086	.022	.054
	β	.062	.073	.028	.057	.033	.074	.016	.067

actual MSE = MSE/100; s.e.: median standard error; rej.f.: rejection frequency.

Table 4. Estimators and t-tests for ρ and β; Design IV-C; 2500 replications.

		N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
estimator	param.	bias	MSE	bias	MSE	bias	MSE	bias	MSE
AB	ρ	-.125	4.013	-.025	.517	-.079	.929	-.018	.099
	β	-.007	.595	-.001	.111	-.014	.187	-.004	.039
ABAS	ρ	-.088	2.982	-.016	.468	-.075	.917	-.010	.078
	β	-.009	.713	-.001	.115	-.015	.249	-.004	.043
OPAS	ρ	-.102	3.008	-.019	.457	-.077	.911	-.013	.080
	β	-.009	.654	-.001	.112	-.014	.211	-.004	.041
SYS	ρ	-.007	.566	.000	.131	-.001	.089	.001	.037
	β	.005	.399	.004	.092	.005	.087	.003	.029
INFLIML	ρ	-.056	.945	-.023	.193	-.020	.152	-.009	.031
	β	.002	.325	-.001	.064	.000	.083	.000	.019
ABLIML	ρ	-.062	1.011	-.024	.204	-.027	.197	-.011	.040
	β	-.010	.442	-.003	.090	-.019	.151	-.005	.033
ABASLIML	ρ	-.063	1.024	-.024	.203	-.027	.197	-.011	.039
	β	-.009	.456	-.003	.090	-.018	.158	-.004	.033
OPASLIML	ρ	-.062	1.017	-.024	.204	-.027	.197	-.011	.039
	β	-.010	.448	-.003	.090	-.018	.153	-.004	.033
ssLIML	ρ	-.076	2.026	-.044	.450	-.032	.324	-.013	.061
	β	-.003	.603	-.007	.122	-.008	.212	-.005	.037
HPTFLIML	ρ	-.037	1.404	-.010	.452	-.016	.257	-.015	.070
	β	-.017	.360	-.018	.096	-.019	.116	-.019	.055
HPT	ρ	.029	3.223	.001	.912	-.017	.354	-.018	.076
	β	-.033	.440	-.031	.160	-.030	.186	-.030	.106
FIML	ρ	.011	1.962	.005	.425	-.001	.251	-.001	.052
	β	-.002	.427	-.001	.084	.000	.129	.000	.027
		s.e.	rej.f.	s.e.	rej.f.	s.e.	rej.f.	s.e.	rej.f.
AB	ρ	.153	.146	.065	.067	.051	.292	.027	.092
	β	.074	.060	.032	.058	.039	.064	.019	.057
SYS	ρ	.071	.053	.032	.044	.027	.056	.019	.066
	β	.061	.060	.028	.057	.026	.057	.017	.048
INFLIML	ρ	.056	.093	.021	.074	.024	.079	.017	.050
	β	.056	.061	.024	.063	.028	.057	.013	.045
ABLIML	ρ	NA	.068	.060	.035	.047	.044	.023	.032
	β	NA	.007	.032	.016	.038	.030	.018	.034
FIML	ρ	.101	.176	.051	.079	.040	.086	.022	.076
	β	.063	.052	.028	.052	.031	.071	.016	.057

actual MSE = MSE/100; s.e.: median standard error; rej.f.: rejection frequency.

Table 5. Estimators for ρ and β; Designs II-C and II'-C; 2500 replications.

		N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
Design II-C									
estimator	param.	bias	MSE	bias	MSE	bias	MSE	bias	MSE
OPASLIML	ρ	-.058	1.087	-.025	.226	-.037	.287	-.012	.051
	β	-.019	.476	-.004	.096	-.028	.217	-.008	.042
HPTFLIML	ρ	-.034	1.479	-.017	.507	-.031	.325	-.026	.119
	β	-.036	.452	-.031	.164	-.034	.214	-.037	.152
Design II′-C									
estimator	param.	bias	MSE	bias	MSE	bias	MSE	bias	MSE
OPASLIML	ρ	-.076	1.400	-.035	.329	-.036	.286	-.013	.051
	β	-.010	.505	-.002	.104	-.030	.230	-.007	.039
HPTFLIML	ρ	-.066	1.394	-.030	.352	-.022	.247	-.009	.052
	β	-.004	.376	-.004	.072	-.011	.111	-.009	.022

actual MSE = MSE/100.

Table 6. Estimators for ρ and β; Designs III-C and III'-C; 2500 replications.

		N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
Design III-C									
estimator	param.	bias	MSE	bias	MSE	bias	MSE	bias	MSE
OPASLIML	ρ	-.058	1.075	-.022	.267	-.025	.263	-.007	.060
	β	-.017	.492	-.005	.101	-.032	.222	-.007	.042
HPTFLIML	ρ	-.045	1.400	-.022	.519	-.026	.304	-.028	.119
	β	-.033	.464	-.033	.169	-.038	.215	-.036	.151
Design III′-C									
estimator	param.	bias	MSE	bias	MSE	bias	MSE	bias	MSE
OPASLIML	ρ	-.063	1.271	-.027	.324	-.030	.278	-.011	.056
	β	-.012	.490	-.003	.104	-.032	.249	-.007	.043
HPTFLIML	ρ	-.057	1.290	-.026	.329	-.018	.237	-.010	.053
	β	-.005	.342	-.005	.070	-.015	.128	-.008	.022

actual MSE = MSE/100.

Table 7. Estimators for ρ and β; Design IV′-C; 2500 replications.

		N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
estimator	param.	bias	MSE	bias	MSE	bias	MSE	bias	MSE
AB	ρ	-.016	.369	-.003	.067	-.018	.120	-.004	.020
	β	.008	.288	.002	.056	.015	.107	.003	.018
SYS	ρ	.002	.262	.005	.054	.003	.065	.005	.017
	β	.007	.265	.005	.053	.007	.069	.006	.018
OPASLIML	ρ	-.018	.220	-.007	.042	-.009	.061	-.003	.012
	β	-.010	.203	.004	.042	.006	.062	.002	.011
HPTFLIML	ρ	-.005	.325	-.006	.066	-.005	.059	-.004	.015
	β	-.010	.241	.010	.059	.005	.056	.005	.013

actual MSE = MSE/100.

of designs II, III and IV when $N = 100$, and design II-D. When the variances of $\underline{\varepsilon}_{i,1}$ and $\underline{\xi}_{i,1}$ in designs II-C and III-C are reduced to the values in designs II′-C and III′-C, then the difference between the MSEs of the FE and RE LIMLEs for ρ decreases or becomes negative and the MSE of the FE LIMLE for β becomes (much) lower than the MSE of the RE LIMLE for β for all sample sizes considered. The last result may be partly due to the fact that in II′-C and III′-C the MSEs of the (preliminary) HPT MLEs for α_x and β_x are lower than the MSEs of the OPAS GMM estimators for α_x and β_x. We also note that if the variances of $\underline{\varepsilon}_{i,1}$ and $\underline{\xi}_{i,1}$ are zero and the RE and FE LIMLE use the same estimator for α_x and β_x, then they are asymptotically equivalent. In this case their MSEs were about the same in unreported MC results.

9 The HPT MLE based FE LIMLEs (often) have a (much) lower MSE than the HPT MLEs except for the FE LIMLE for β in design IV-B when $N = 100$.

10 When the data become more persistent, we find that in all four designs, i.e., in I-IV the MSEs of our RE and FE LIMLEs increase. When $N = 100$, the MSEs of the RE LIMLEs increase faster when moving from the B versions to the C versions of designs I-IV than the MSEs of the FE LIMLE, whereas when $N = 500$ and $T = 10$ and for designs I and III also when $N = 500$ and $T = 5$, we find the opposite.

11 The ABLIML based Wald test has (much) better size properties than the AB GMM based Wald test, especially when N is small (e.g., $N = 100$).

12 When $T = 5$ and $N = 100$, computation of the Murphy and Topel standard errors for the ABLIMLE sometimes failed and therefore we omit results for this case. If that happened, the corresponding Wald test did not reject the null hypothesis. Alternatively, one can use a simple nonparametric bootstrap to conduct inference.

Summarizing, the RE LIMLE usually performs (much) better than the RE FIMLE, the ss-LIMLE and the FE LIMLE, and the latter normally performs (much) better than the HPT MLE. However, when mean stationarity is likely to hold or the data are not close to normally distributed, a System GMM or Ahn-Schmidt type GMM estimator may be preferable to the RE LIQMLE.

Table 8. Estimators for ρ and β; Design IV″-E; 2500 replications.

		N=100, T=5		N=500, T=5		N=100, T=10		N=500, T=10	
estimator	param.	bias	MSE	bias	MSE	bias	MSE	bias	MSE
AB	ρ	-.031	.814	-.006	.129	-.027	.198	-.006	.029
	β	.007	.534	.001	.097	.005	.181	.001	.034
SYS	ρ	.063	.514	.067	.475	.071	.532	.074	.558
	β	.030	.500	.028	.152	.035	.254	.039	.179
OPASLIML	ρ	-.027	.372	-.010	.074	-.009	.077	-.003	.016
	β	.004	.378	.001	.080	.000	.135	.000	.026
HPTFLIML	ρ	.025	1.405	.004	.193	.001	.124	.000	.022
	β	.012	.422	.010	.088	.008	.134	.008	.030

actual MSE = MSE/100.

4. EMPIRICAL ILLUSTRATION: EMPLOYMENT AND WAGES

We revisit the application discussed in Alonso-Borrego and Arellano (1999, henceforth AA) and Arellano (2003). Using a balanced panel of 738 Spanish manufacturing firms with annual data for the period 1983-1990, they estimated various versions of a bivariate VAR(2) model for the logarithms of employment and real wages, denoted $n_{i,t}$ and $wa_{i,t}$ respectively. It nests a model that can be regarded as the reduced form of an intertemporal model of employment determination under rational expectations. Their estimators include the two-step optimal AB and System GMM estimators and their Symmetrically Normalised GMM and LIML (or continuously updated, CUE) analogues. Additive individual and time effects were allowed for in both equations of the model. Wages were measured by dividing the total wage bill of the firm by employment. See AA and Arellano (2003, p. 116ff.) for a detailed description of the data, models and estimators.

The wage equation in AA only includes its own lags whereas Arellano (2003) estimated an unrestricted VAR(2) model. However, the estimates of the coefficients of the two lags of employment in his wage equation became significant only when he used the System estimator. At any rate, as wages are measured at the firm level, one cannot rule out the possibility that lags of employment affect wages. Therefore we too estimated the unrestricted VAR model. Like Arellano (2003), we did not impose restrictions on the parameters that rule out explosive behaviour. Furthermore, like Arellano (2003), we treated wage and employment as predetermined regressors in both equations of the VAR and allowed the regressors to be correlated with the unobserved individual effects. We estimated the VAR model using the multi-step optimal ABAS and OPAS (GMM) estimators and the RE LIQMLEs based on them (i.e., the RE LIQMLEs that use them to construct the residuals that serve as controls for endogeneity), the FE LIQMLE based on a FE MLE of HPT (2002) as well as two-step optimal AB, System and AS (GMM) estimators. The first six estimators have been computed using our own GAUSS program. The last three estimators have been computed using the xtdpdgmm program in STATA.[19] The estimates produced by GAUSS and STATA are reported in tables 9 and 10, respectively.

Our GMM estimators use the full sets of available moment conditions. For instance, the AB estimators use second and all available higher order lags of employment and wages as instruments. We note that the STATA results related to the AB and System estimators are identical to those reported in Arellano (2003). We considered three versions of the LIQMLEs, which correspond to Models 1-3. The first version imposes homoskedaticity over time, the second version allows for different variance parameters over time, while the third version is the fully heteroskedastic version which also replaces the individual effect by an equation similar to (2.9). For the AB and System estimators we report conventional, bootstrap and Windmeijer's (2005) robust standard errors, for the AS-type estimators we only report the first two kinds of standard errors, while for the LIQMLEs we only report the (nonparametric) bootstrap standard errors, which are based on resampling the individual data vectors.[20] In the case of the LIQMLEs we also report (the average of) the estimates of the coefficients of the added residuals (the controls) and (the average of) their standard errors. Finally, table 10 includes results for Sargan-Hansen (J) tests of

[19]The AB estimates obtained by using the GAUSS and STATA programs are slightly different. The OPAS and AS estimates are also slightly different.

[20]STATA's xtdpdgmm program also computes Windmeijer's (2005) robust standard errors for the AS estimator but in our case we found that they were smaller than conventional first-order asymptotic standard errors, which should not be the case.

overidentifying restrictions and in the context of the AS estimates also results for tests of lack of first- and second-order serial correlation of the errors. Table 10 also reports J-test results corresponding to GMM estimation with 'collapsed' instruments.

Table 9. VAR estimates for Employment and Wage using Spanish panel data and GAUSS.

	employment			wage								
estimator	est.	r.s.e.	b.s.e.	est.	r.s.e.	b.s.e.						
AB												
$n_{i,t-1}$	0.86	.128	.144	-0.02	.119	.104						
$n_{i,t-2}$	0.00	.050	.058	0.03	.055	.055						
$wa_{i,t-1}$	0.05	.117	.130	0.27	.140	.137						
$wa_{i,t-2}$	-0.06	.051	.053	0.02	.038	.043						
ABAS												
$n_{i,t-1}$	1.12	NA	.116	0.02	NA	.109						
$n_{i,t-2}$	-0.09	NA	.050	-0.02	NA	.040						
$wa_{i,t-1}$	0.26	NA	.124	0.70	NA	.100						
$wa_{i,t-2}$	-0.06	NA	.028	0.07	NA	.037						
OPAS												
$n_{i,t-1}$	0.99	NA	.113	-0.03	NA	.101						
$n_{i,t-2}$	-0.04	NA	.047	0.01	NA	.041						
$wa_{i,t-1}$	0.15	NA	.106	0.51	NA	.104						
$wa_{i,t-2}$	-0.06	NA	.028	0.05	NA	.035						
	Model 1				Model 2				Model 3			
	employment		wage		employment		wage		employment		wage	
	est.	b.s.e.	est.	b.s.e.	est.	b.s.e.	est.	b.s.e.	est.	b.s.e.	est.	b.s.e.
ABASLIML												
$n_{i,t-1}$	1.12	.034	0.18	.043	1.18	.043	0.19	.056	1.18	.053	0.19	.065
$n_{i,t-2}$	-0.08	.032	-0.07	.042	-0.11	.034	-0.08	.043	-0.11	.040	-0.09	.050
$wa_{i,t-1}$	0.25	.053	0.70	.058	0.26	.054	0.74	.068	0.28	.057	0.74	.070
$wa_{i,t-2}$	-0.04	.038	0.11	.043	-0.05	.035	0.11	.044	-0.05	.038	0.11	.043
$Resid_{i(,t)}$	-0.09	.031	-0.15	.040	-0.09	.025	-0.14	.040	-0.10	.036	-0.15	.060
OPASLIML												
$n_{i,t-1}$	1.13	.034	0.20	.041	1.19	.038	0.22	.054	1.20	.048	0.23	.054
$n_{i,t-2}$	-0.08	.032	-0.06	.042	-0.11	.034	-0.08	.042	-0.12	.040	-0.08	.052
$wa_{i,t-1}$	0.27	.052	0.76	.051	0.29	.051	0.81	.061	0.31	.051	0.81	.066
$wa_{i,t-2}$	-0.02	.037	0.14	.040	-0.04	.036	0.14	.041	-0.03	.038	0.13	.055
$Resid_{i(,t)}$	-0.07	.023	-0.12	.031	-0.07	.018	-0.11	.040	-0.08	.030	-0.12	.060
HPTFLIML												
$n_{i,t-1}$	1.11	.028	0.15	.027	1.16	.037	0.16	.033	1.16	.039	0.18	.040
$n_{i,t-2}$	-0.07	.033	-0.06	.032	-0.10	.035	-0.08	.038	-0.10	.037	-0.09	.042
$wa_{i,t-1}$	0.22	.031	0.77	.031	0.24	.035	0.82	.038	0.25	.041	0.82	.040
$wa_{i,t-2}$	-0.06	.029	0.15	.034	-0.07	.031	0.15	.032	-0.07	.032	0.15	.037
$Resid_{i(,t)}$	-0.11	.009	-0.12	.008	-0.11	.008	-0.11	.015	-0.11	.019	-0.11	.032

est.: parameter estimate; r.s.e.: robust Windmeijer (2005) standard error for AB estimator; b.s.e. bootstrap standard error for all estimators.

None of the J-tests reject the validity of the various sets of moment conditions at a significance level of 5%, although the J-tests associated with the System estimator have low p-values. Nonetheless, a Sargan Difference test rejects the validity of the extra moment conditions that are exploited by the AS estimator for the wage equation but not by the AB estimator for that equation: $J_{\text{AS,wage}} - J_{\text{AB,wage}} = 16.94 > \chi^2_{0.95}(4) = 9.49$. However, the p-value for $J_{\text{AB,wage}}$ is very high (0.974), while the p-value for $J_{\text{AS,wage}}$ is a very reasonable 0.545. A very high p-value for a J-test may reflect size-distortions that occur when the number of overidentifying restrictions is high. If we re-estimate the model using 'collapsed' instruments, we obtain $\hat{J}_{\text{AS,wage}} - \hat{J}_{\text{AB,wage}} = 8.61 < \chi^2_{0.95}(4) = 9.49$ so the extra moment conditions exploited by the AS estimator for the wage equation are probably valid after all. The Sargan Difference tests reject mean stationarity for both equations of the VAR model: $J_{\text{SYS,emp}} - J_{\text{AS,emp}} = 16.28 > \chi^2_{0.95}(8) = 15.507$ and $J_{\text{SYS,wage}} - J_{\text{AS,wage}} = 25.87 > 15.507$ (but $\hat{J}_{\text{SYS,emp}} - \hat{J}_{\text{AS,emp}} = 13.24 < \chi^2_{0.95}(8) = 15.507$, while $\hat{J}_{\text{SYS,wage}} - \hat{J}_{\text{AS,wage}} = 15.514 > 15.507$). These results combined with the low p-values associated with $J_{\text{SYS,emp}}$ and $J_{\text{SYS,wage}}$ cast serious doubt on the validity of the moment conditions that rely on mean stationarity and hence the System GMM estimates, at least in the case of the wage equation. The LIQML estimates are very similar across the three models. Note also that their standard errors are mostly smaller than those of the ABAS and OPAS estimates. The OPAS estimates lie in between the AB and ABAS estimates. The

Table 10. VAR estimates for Employment and Wage using Spanish panel data and STATA.

	employment			wage		
AB	est.	conv. s.e.	robust s.e.	est.	conv. s.e.	robust s.e.
$n_{i,t-1}$	0.842	0.088	0.128	-0.042	0.100	0.119
$n_{i,t-2}$	-0.003	0.029	0.050	0.050	0.034	0.055
$wa_{i,t-1}$	0.078	0.084	0.117	0.256	0.109	0.140
$wa_{i,t-2}$	-0.053	0.025	0.051	0.025	0.025	0.038
	J		$(\hat{J} - collapsed)$	J		$(\hat{J} - collapsed)$
	36.91		5.75	21.40		5.80
(d.f.)	36		8	36		8
p-value	0.427		0.675	0.974		0.669
SYS	est.	conv. s.e.	robust s.e.	est.	conv. s.e.	robust s.e.
$n_{i,t-1}$	1.170	0.026	0.047	0.079	0.027	0.047
$n_{i,t-2}$	-0.126	0.019	0.033	-0.063	0.021	0.029
$wa_{i,t-1}$	0.132	0.021	0.027	0.776	0.022	0.034
$wa_{i,t-2}$	-0.110	0.021	0.029	0.084	0.023	0.034
	J		$(\hat{J} - collapsed)$	J		$(\hat{J} - collapsed)$
	61.21		25.99	64.21		29.92
(d.f.)	48		20	48		20
p-value	0.096		0.166	0.059		0.071
AS	est.	conv. s.e.	robust s.e.	est.	conv. s.e.	robust s.e.
$n_{i,t-1}$	0.973	0.092	NA	-0.061	0.109	NA
$n_{i,t-2}$	-0.036	0.042	NA	0.018	0.038	NA
$wa_{i,t-1}$	0.147	0.120	NA	0.514	0.085	NA
$wa_{i,t-2}$	-0.056	0.024	NA	0.064	0.033	NA
	J		$(\hat{J} - collapsed)$	J		$(\hat{J} - collapsed)$
	44.93		12.75	38.34		14.41
(d.f.)	40		12	40		12
p-value	0.273		0.387	0.545		0.275
	$m1$	$m2$		$m1$	$m2$	
	-9.45	0.75		-9.09	0.16	
p-value	0.000	0.456		0.000	0.875	

est.: parameter estimates based on full (i.e., non-collapsed) sets of instruments; conv. s.e.: conventional (i.e., first-order asymptotic) standard error; robust s.e.: Windmeijer (2005) s.e.; $\hat{J} - collapsed.$: Sargan-Hansen test for model with 'collapsed' instruments; AS is a two-step Ahn-Schmidt type estimator.

ABAS estimates are similar to the LIQML estimates with the exception of the estimates of the coefficients of the lags of employment in the wage equation. The ABAS (and OPAS) estimates of the latter are insignificant, while the LIQML estimates of the coefficient of the first lag of employment in the wage equation are strongly significant implying that wage should not be treated as strictly exogenous in the employment equation. Similarly, employment is not strictly exogenous in the wage equation. Furthermore, both variables are not weakly exogenous because the added residuals are significant in both equations. Concluding, our LIQML estimates are not very different from the System estimates in Arellano (2003), which, however, rely on the assumption of mean stationarity that is rejected by various tests. In particular, like the System estimation results, our LIQML estimation results show that the log of wage is affected by lagged levels of the log of employment.

5. CONCLUDING REMARKS

In this paper we discussed large N, fixed T consistent RE and FE limited and full information Quasi ML estimators for variations of the conditional panel AR(1) model with one additional regressor under alternative exogeneity assumptions about the regressor. As mentioned in the introduction, these methods can be extended straightforwardly to models with multiple lags and additional regressors. We first proposed LIQMLEs for models with a predetermined regressor that can be correlated with the individual effect(s). They are based on a two-step control function approach. If there is more than one additional regressor, the RE LIQMLE is more robust than Bai's (2013b) RE FIQMLE because it remains consistent when the model of a regressor that is weakly exogenous w.r.t. ρ and β is misspecified, whereas the FIQMLE can become inconsistent in that case. The LIQMLEs are also more easily computed than the FIQMLEs especially when the system is not small. We then generalized these LIMLEs to allow for endogeneity of the regressor by adding the composite residual of the equation for $x_{i,t}$ or $\widetilde{\Delta} x_{i,t}$ as a regressor to the equation for $y_{i,t}$ or $\widetilde{\Delta} y_{i,t}$ for each $t \in \{2, \ldots, T\}$.

Under normally distributed data, GMM estimators for a model for some variable y that contains a lag of y and a predetermined regressor x that is not weakly exogenous are less precise but more robust than our two-step LIMLEs when the GMM estimators do not exploit moment conditions that are (partly) based on a correctly specified model for x. Nevertheless, if one considers using lags of x as instruments for the model for y, one still needs to know the order of the lowest lag of y on which x depends.

When ρ in (1.1) or (1.2) equals unity and $x_{i,t}$ and μ_i drop out of the model, then 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 a slower rate of convergence and non-standard distributions. Kruiniger (2016) discusses Quasi LM tests for hypotheses about ρ that have asymptotically correct size in a uniform sense.

In case of cross-sectional heteroskedasticity, one may be able to obtain more precise QMLEs by modelling the conditional variances of the loadings and the idiosyncratic errors, e.g. as functions of the initial observations for y and x.

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).

REFERENCES

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.

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.

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

Alonso-Borrego, C., and M. Arellano, 1999, Symmetrically Normalized Instrumental-Variable Estimation Using Panel data, Journal of business & economic statistics 17, 36-49.

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

Alvarez, J. and M. Arellano, 2004, Robust likelihood estimation of dynamic panel data models, unpublished manuscript, CEMFI, Madrid, Spain.

Anderson, T.W., and C. Hsiao, 1981, Estimation of dynamic models with error components, Journal of the American Statistical Associasion 76, 598-606.

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

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.

Arellano, M., 2003, Panel Data Econometrics, Oxford University Press.

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.

Arellano, M., and O. Bover, 1995, Another look at the instrumental variable estimation of error-components models, Journal of econometrics 68, 29-51.

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

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

Bai, J. and K. Li, 2012, Statistical analysis of factor models of high dimension, The Annals of Statistics 40, 436-465.

Bhargava, A. and J.D. Sargan, 1983, Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods, Econometrica 51, 1635-1659.

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.

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

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

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

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.

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.

Juodis, A., V. Sarafidis, 2018, Fixed T dynamic panel data estimators with multifactor errors, Econometric reviews 37, 893-929.

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

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

Kruiniger, H., 2016, Uniform Quasi ML based inference for the panel AR(1) model with arbitrary initial conditions, unpublished manuscript.

Kruiniger, H., 2019, Common correlated effects estimation of dynamic panel data models for short panels, unpublished manuscript.

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

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

Murphy, K.M, and R.H. Topel, 1985, Estimation and Inference in Two-Step Econometric Models, Journal of Business & Economic Statistics 3, 88-97.

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.

Oberhofer, W. and Kmenta, J., 1974, A general procedure for obtaining maximum likelihood estimates in generalized regression models, Econometrica 42, 579-590.

Pesaran, M.H., 2006, Estimation and inference in large heterogeneous panels with a multifactor error structure, Econometrica 74, 967-1012.

Williams, R., P.D. Allison and E. Moral-Benito, 2017, xtdpdml: Linear Dynamic Panel Data Estimation using Maximum Likelihood and Structural Equation Modeling, in progress, last revised 17 September 2017, University of Notre Dame, University of Pennsylvania and Banco de Espana.

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

Wooldridge, J.M., 2010, Econometric Analysis of Cross Section and Panel Data, second edition, MIT Press, Cambridge MA.