Dynamic Panel Data Modeling using Maximum Likelihood: An Alternative to Arellano-Bond*

Enrique Moral-Benito
Banco de España

Paul D. Allison
University of Pennsylvania

Richard Williams
University of Notre Dame

February 13, 2018

Abstract

The Arellano and Bond (1991) estimator is widely-used among applied researchers when estimating dynamic panels with fixed effects and predetermined regressors. This estimator might behave poorly in finite samples when the cross-section dimension of the data is small (i.e. small N), especially if the variables under analysis are persistent over time. This paper discusses a maximum likelihood estimator that is asymptotically equivalent to Arellano and Bond (1991) but presents better finite sample behavior. The estimator is based on an alternative parametrization of the likelihood function introduced in Moral-Benito (2013). Moreover, it is easy to implement in Stata using the `xtdpdml` command as described in the companion paper Williams et al. (2018), which also discusses further advantages of the proposed estimator for practitioners.

JEL Codes: C23.
Keywords: dynamic panel data, maximum likelihood estimation.

*The authors thank valuable comments by Manuel Arellano, Kristin MacDonald, an anonymous referee, and attendants to seminars held at Bank of Spain, the 2016 Spanish Stata Users Group meeting in Barcelona, and the 2015 Stata Users Conference in Columbus, Ohio. Code and data used in this article are available on the website https://www3.nd.edu/~rwilliam/dynamic/index.html which includes further materials related to the practical implementation of the estimator.

1 Introduction

Panel data are very popular among applied researchers in many different fields from economics to sociology. A panel data set is one that follows a given sample of subjects over time, and thus provides multiple observations on each subject in the sample. Subjects may be workers, countries, firms, regions... while the multiple observations per subject usually refer to different moments in time (e.g. years, quarters, or months). Indeed, time series and cross-sectional data can be thought of as special cases of panel data that are in one dimension only (one panel subject for the former, one time point for the latter).

Allowing for the presence of subject-specific unobserved heterogeneity represents one of the key advantages of using panel data. Having multiple observations per individual allows identifying a time invariant component that is unobserved to the econometrician and may be correlated with other observable characteristics in the data set. For instance, in cross-country studies of economic growth, unobserved heterogeneity at the country level may be associated with cultural differences or geographical characteristics across countries (see Islam, 1995). Moreover, in a regression of y on x, panel data can accommodate feedback effects from current y to future x, so that this particular form of reverse causality can easily be accounted for by using well-known panel data techniques where the x regressors are said to be predetermined (see Chapter 8 in Arellano, 2003).[1] Predetermined regressors are also labeled as weakly exogenous or sequentially exogenous in the literature (Wooldridge, 2010). Dynamic panels in which the regressors include the lagged dependent variable are the best example in this category. This is so because feedback from current y to future y exists by construction (see for instance Arellano and Bond, 1991).

The panel GMM estimator discussed in Arellano and Bond (1991) is probably the most popular alternative for estimating dynamic panels with unobserved heterogeneity and predetermined regressors. To be more concrete, the typical model to be estimated is given by the traditional partial adjustment with feedback model, which is very popular among economists (see Arellano (2003) page 143). The beauty of the Arellano and Bond (1991) estimator is that relies on minimal assumptions and provides consistent estimates even in panels with few time series observations per individual (i.e. small T). However, it does require large samples in the cross-section dimension (i.e. large N) and

[1] Intuitively, this assumption implies that only future values of the explanatory variables are affected by the current value of the dependent variable.

its finite sample performance might represent a concern when the number of units in the panel is relatively small, especially if the variables under analysis are persistent (see Moral-Benito, 2013).

Against this background, several alternative estimators have been proposed in the literature with the same identifying assumption. For instance, Alonso-Borrego and Arellano (1999), Ahn and Schmidt (1995), and Hansen et al. (1996) consider different GMM variants of the Arellano and Bond (1991) estimator with better finite sample performance. Also, likelihood-based approaches have been considered under similar identifying assumptions resulting in better finite sample behavior (e.g. Hsiao et al. (2002), Moral-Benito (2013)). A practical limitation of these alternatives is that their implementation by practitioners is far from straightforward given the requirement of certain programming capabilities as well as numerical optimization routines.

We do not include in the above category the so-called system-GMM estimator by Arellano and Bover (1995) and Blundell and Bond (1998) because it requires an additional identifying assumption for consistency. In particular, it relies on the mean stationarity assumption that has been proved to be controversial in most empirical settings. Intuitively, this assumption requires that the variables observed in the data set come from dynamic processes that started in the distant past so that the have already reached their steady state distribution, which is hard to motivate in panels of young workers or firms as well as country panels starting just after WWII (see Barro and Sala-i-Martin, 2003). On the other hand, as pointed out by Bazzi and Clemens (2013), concern has intensified in recent years that many instrumental variables of the type considered in panel GMM estimators such as Arellano and Bond (1991) and Arellano and Bover (1995) may be invalid, weak, or both. The effects of this concern may be substantial in practice as recently illustrated by Kraay (2015).

In this paper, we discuss a maximum likelihood estimator based on the same identification assumption as Arellano and Bond (1991) so that both alternatives are asymptotically equivalent. However, we show in Section 3 that our likelihood-based alternative is strongly preferred in terms of finite sample performance, especially when the number of units in the panel (N) is small. Moreover, as illustrated in some of our simulations as well as in Williams et al. (2018), there are situations in which the likelihood approach may be preferred to standard GMM even when N is large and the unbalancedness represents a concern.

The particular likelihood function presented in this paper is an alternative parametrization to the one presented in Moral-Benito (2013) but based on the same set of assumptions. In particular, it can be interpreted as an intermediate situation between the full covariance structure (FCS) and the simultaneous equation model (SEM) representation discussed in Moral-Benito (2013). This is so because the restrictions are enforced in the covariance matrix as in the SEM representation, but the analysis is not conditional on the initial observations as in the FCS parametrization (see also Allison, 2005; Allison et al. 2017).

This particular likelihood is useful in practice because it can be maximized using numerical optimization techniques available in standard software packages. To be more concrete, the maximum likelihood estimator discussed in this paper is easy to implement in Stata adapting the `sem` command as described in the companion paper by Williams et al. (2018). The intuition is that period-by-period equations from the panel data model are used to form a system of equations of the type considered in SEM models (see e.g. Bentler and Weeks, 1980). Moreover, there are other software packages that can estimate this model by maximum likelihood including LISREL, EQS, Amos, Mplus, PROC CALIS (in SAS), lavaan (for R), and OpenMx (for R).

The rest of the paper is organized as follows. Section 2 describes the likelihood function. Section 3 illustrates the finite sample performance of the proposed estimator in comparison to the Arellano and Bond (1991) GMM alternative. In Section 4 we illustrate the usefulness of the estimator in the context of an empirical application investigating the effect of financial development on economic growth across countries based on Levine et al. (2000). Section 5 concludes.

2 Partial Adjustment with Feedback

We consider the following model:

$$y_{it} = \lambda y_{it-1} + \beta x_{it} + \alpha_i + v_{it} \tag{1}$$

$$E\left(v_{it} \mid y_i^{t-1}, x_i^t, \alpha_i\right) = 0 \qquad (t = 1, ..., T)(i = 1, ..., N) \tag{2}$$

where i indexes units in the panel (workers, countries, firms...) and t refers to time periods (decades, years, quarters...). We also define the $t \times 1$ vectors of past realizations $y_i^{t-1} = (y_{i,0}, ..., y_{i,t-1})'$ and $x_i^t = (x_{i,1}, ..., x_{i,t})'$. Note that β and x_{it} can also be vectors including more than one predetermined regressor. In addition, we can easily include strictly exogenous regressors.

This model relaxes the strict exogeneity assumption for the x variables. The assumption in (1) allows for feedback from lagged values of y to the current value for x. Moreover it implies lack of autocorrelation in v_{it} since lagged vs are linear combinations of the variables in the conditioning set. Crucially, assumption (2) is the only assumption we impose throughout the paper.[2] Indeed, this is also the only assumption required for consistency of the Arellano and Bond (1991) GMM estimator.

Time invariant regressors can also be included, under the assumption that they are uncorrelated with the fixed effects, and advantage over the Arelllano and Bond (1991) approach. Finally, in addition to the individual-specific effects α_i, we can allow cross-sectional dependence by including a set of time dummies. However, for the sake of exposition we focus on specification (1) that features the main ingredients of the approach and facilitates its illustration.

2.1 The Likelihood Function

In the spirit of Allison (2005) and Allison et al. (2017), this section develops a parameterization of the model in (1)-(2) that leads to a maximum likelihood estimator that is asymptotically equivalent to the Arellano and Bond (1991) estimator augmented with the moment condition arising from lack of autocorrelation as discussed in Ahn and Schmidt (1995). Moral-Benito (2013) also consider alternative parametrizations of the same model. In particular, the restrictions implied by (2) can be placed in either the coefficient matrices or the variance-covariance matrix depending on how the system of equations is written. The parametrization considered here is useful because it can be easily implemented in practice using the `sem` command in Stata as described in Williams et al. (2018). Note also that other SEM packages such as Mplus, PROC CALIS in SAS, and lavaan or OpenMx in R can also be used.

In addition to the T equations given by (1), we complete the model with an equation for y_{i0} as well as T additional reduced-form equations for x:[3]

[2]Despite we derive the log likelihood under normality, it is important to remark that the resulting estimator is consistent and asymptotically normal regardless of non-normality.

[3]Needless to say, additional x predetermined regressors can be included as well as other exogenous covariates. We only discuss this canonical specification for the sake of notation simplicity.

$$
\begin{aligned}
y_{i0} &= v_{i0} && (3)\\
x_{i1} &= \xi_{i1} && (4)\\
&\vdots \\
x_{iT} &= \xi_{iT} && (5)
\end{aligned}
$$

In order to rewrite the system of equations given by (1) and (3)-(5) in matrix form, we define the following vectors of observed data (R_i) and disturbances (U_i):

$$
\begin{aligned}
R_i &= (y_{i1}, ..., y_{iT}, y_{i0}, x_{i1}, ...x_{iT})' && (6)\\
U_i &= (\alpha_i, v_{i1}, ..., v_{iT}, v_{i0}, \xi_{i1}, ...\xi_{iT})' && (7)
\end{aligned}
$$

Importantly, the covariance matrix of the disturbances captures the restrictions imposed by (2) and it is given by:

$$
Var\,(U_i) = \Sigma = \left(\begin{array}{c|c} \Sigma_{11} & \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right) = \left(\begin{array}{cccc|cccc}
\sigma_\alpha^2 & & & & & & & \\
0 & \sigma_{v1}^2 & & & & & & \\
\vdots & \vdots & \ddots & & & & & \\
0 & 0 & \cdots & \sigma_{vT}^2 & & & & \\
\hline
\phi_0 & 0 & \cdots & 0 & \sigma_{v0}^2 & & & \\
\phi_1 & 0 & \cdots & 0 & \omega_{01} & \sigma_{\xi_1}^2 & & \\
\phi_2 & \psi_{21} & \cdots & 0 & \omega_{02} & \omega_{12} & \sigma_{\xi_2}^2 & \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \\
\phi_T & \psi_{T1} & \psi_{T2} & \cdots & \omega_{0T} & \omega_{1T} & \cdots & \sigma_{\xi_T}^2
\end{array}\right) \qquad (8)
$$

where the element Σ_{21} captures the correlation between the fixed effects and the regressors through the ϕ parameters, and the feedback process from y to x allowing for nonzero correlations between the current vs and future ξs:

$$
cov(v_{ih}, \xi_{it}) = \begin{cases} \psi_{th} & \text{if } h < t \\ \mathbf{0} & \text{otherwise} \end{cases} \qquad (9)
$$

On the other hand, Σ_{11} gathers the lack of autocorrelation in the v disturbances and the fixed effects α_i, and Σ_{22} gathers all of the contemporaneous and dynamic relationships between the x variables. In contrast to the standard Arellano and Bond (1991) approach, we can accommodate time-varying error variances in Σ_{11}.

Note that the covariance matrix of the joint distribution of the initial observations (y_{i0}, x_{i1}) and the individual effects α_i is unrestricted with the corresponding covariances captured through the parameters ϕ_0, ϕ_1, and ω_{01}. This is in sharp contrast with the mean stationarity assumption required by the so-called system-GMM estimator discussed in Arellano and Bover (1995) and Blundell and Bond (1998).

We next define the following matrices of coefficients:

$$B = \left(\begin{array}{ccccc|ccccc} 1 & 0 & 0 & \cdots & 0 & -\lambda & -\beta & 0 & \cdots & 0 \\ -\lambda & 1 & 0 & \cdots & 0 & 0 & 0 & -\beta & \cdots & 0 \\ 0 & -\lambda & 1 & \cdots & 0 & 0 & \vdots & & \ddots & \\ \vdots & & & \ddots & & \vdots & & & & \\ 0 & \cdots & & -\lambda & 1 & 0 & 0 & \cdots & 0 & -\beta \\ \hline 0 & & \cdots & & 0 & & & & & \\ \vdots & & \ddots & & \vdots & & & I_{T+1} & & \\ 0 & & \cdots & & 0 & & & & & \end{array}\right)$$

$$D = \left(\begin{array}{c|c} d & I_{2T+1} \end{array}\right)$$

where $d = (1, ..., 1, 0, ..., 0)'$ is a column vector with T ones and $T+1$ zeros.

We can now write equations (1) and (3)-(5) in matrix form:

$$BR_i = DU_i \tag{10}$$

Thus, assuming normality, the joint distribution of R_i is:

$$R_i \sim N\left(0, B^{-1}D\Sigma D'B'^{-1}\right) \tag{11}$$

with resulting log-likelihood:

$$L \propto -\frac{N}{2}\log\det\left(B^{-1}D\Sigma D'B'^{-1}\right) - \frac{1}{2}\sum_{i=1}^{N} R_i'\left(B^{-1}D\Sigma D'B'^{-1}\right)^{-1} R_i \tag{12}$$

As shown by Moral-Benito (2013), the maximizer of L is asymptotically equivalent to the Arellano and Bond (1991) GMM estimator[4] regardless of non-normality. In Appendix A we illustrate, for the

[4] To be more concrete, the asymptotic equivalence is only guaranteed if we augment the Arellano and Bond (1991) estimator with moments resulting from lack of autocorrelation in the errors as discussed by Ahn and Schmidt (1995).

case of $T = 3$ that the number of over-identifying restrictions is the same in both cases. Also, it is worth highlighting that likelihood ratio tests of the model's over-identifying restrictions can be used to test these and other hypotheses of interest.

The likelihood function in equation (12) is derived for balanced panels, i.e., panels in which there are non-missing values for all variables and all individuals at all time periods.[5] However, unbalanced panels are very common in practice. The simplest approach for considering the ML estimator in unbalanced panels is based on the so-called listwise deletion, which is based on eliminating those individuals that have missing values in any of the variables included in the model. This alternative may perform poorly under heavily unbalanced data because the cross-section dimension (N) is drastically reduced generating convergence failures of the likelihood maximization procedure.

Alternatively, we consider the FIML approach discussed in Arbuckle (1996) in order to implement our ML estimator under unbalanced panels. This approach computes individual-specific contributions to the likelihood function using only those time periods that are observed for each individual. Then, the likelihood function to be maximized is computed by accumulating all the individual-specific likelihoods. This alternative has been shown to perform much better than listwise deletion in cross-sectional settings (see Enders and Bandalos, 2001). Indeed, in Section 3 below, we illustrate that the method performs relatively well when working with unbalanced panels using the FIML approach.

3 Simulation Results

In this section, we explore the finite sample behavior of the likelihood-based estimator discussed in this paper in comparison with the Arellano and Bond (1991) GMM estimator.[6] For this purpose, we consider the simulation setting in Bun and Kiviet (2006) also considered by Moral-Benito (2013).

To be more concrete, the data for the dependent variable y and the explanatory variable x are generated according to:

$$y_{it} = \lambda y_{it-1} + \beta x_{it} + \alpha_i + v_{it} \tag{13}$$

$$x_{it} = \rho x_{it-1} + \phi y_{it-1} + \pi \alpha_i + \xi_{it} \tag{14}$$

[5] Note that the GMM approach in Arellano and Bond (1991) can easily handle unbalanced panels by using all information available.

[6] We use the `xtdpdml` Stata command for the maximum likelihood estimator and the `xtdpd` Stata command for the Arellano and Bond (1991) GMM estimator.

where v_{it}, ξ_{it}, and α_i are generated as $v_{it} \sim i.i.d.(0,1)$, $\xi_{it} \sim i.i.d.(0,6.58)$, and $\alpha_i \sim i.i.d.(0,2.96)$. The parameter ϕ in (14) captures the feedback from the lagged dependent variable to the regressor. This particular DGP corresponds to scheme 2 in Bun and Kiviet (2006), which is more realistic than their baseline scheme 1, considered for convenience in the evaluation of their analytical results. With respect to the parameter values, we follow the baseline Design 5 in Moral-Benito (2013) and we fix $\lambda = 0.75$, $\beta = 0.25$, $\rho = 0.5$, $\phi = -0.17$, and $\pi = 0.67$. This configuration allows for fixed effects correlated with the regressor as well as feedback from y to x. Bun and Kiviet (2006) provide more details about this particular Data-Generating Process.

The finite sample performance of the likelihood-based estimator discussed in this paper is compared with the widely-used Arellano and Bond (1991) GMM estimator. Our main motivation is to illustrate the potential gains in terms of finite sample biases of using our maximum likelihood estimator as an alternative to the Arellano and Bond (1991) approach.

Table 1 presents the simulation results. Columns (1) and (2) illustrate that our maximum likelihood estimator (henceforth ML) presents much lower biases when estimating λ than the Arellano and Bond (1991) estimator (henceforth AB) as long as N is small. In the case $T = 4$, the ML bias is negligible even with $N = 100$ while the AB bias is non-negligible (around 5%) even with $N = 1,000$. Turning to β in columns (3) and (4), the same pattern arises with a bias above 7% in the AB estimator when $N = 1,000$. This result points to a significantly better finite sample performance of the ML estimator when the cross-section dimension is small. Not surprisingly, the performance of the AB estimator improves as N increases; therefore, when working with sample sizes around 5,000 individuals or more, the gains from using the ML estimator are relatively minor. The bottom rows of Table 1 investigate the effect of increasing T, the time series dimension of the panel, when N is small. Overall, the performance of the AB estimator improves as T increases while that of the ML estimator remains virtually unaffected. In any case, as long as N is small (e.g. $N = 100$), the ML estimator appears to be preferred to the AB alternative in terms of finite sample biases.

With respect to efficiency, the ML estimator presents lower interquartile ranges for all sample sizes when $T = 4$ as shown in columns (5)-(8). Indeed, the ML estimator is asymptotically efficient under normality as $N \to \infty$. Only in some cases when T increases for N fixed the ML iqrs are slightly larger than those of AB (see the rows $N = 100$, $T = 8$ and $N = 100$, $T = 12$). However, when looking at the root mean square errors (RMSE) in columns (9)-(12), ML presents always lower

RMSEs than AB for λ, and virtually equal for β as T increases.

Finally, when both N and T are relatively large ($N = 5000$, $T = 12$) as in the last row of Table 1, AB and ML perform similarly with negligible biases and low interquartile ranges in both cases.

Table 1: Simulation results.

	Bias λ		Bias β		iqr λ		iqr β		RMSE λ		RMSE β	
	AB	ML	AB	ML	AB	ML	AB	ML	AB	ML	AB	ML
Sample size	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$N = 100$, $T = 4$	-0.220	-0.009	-0.087	0.001	0.389	0.203	0.169	0.115	0.375	0.159	0.158	0.090
$N = 200$, $T = 4$	-0.138	-0.002	-0.054	0.002	0.312	0.167	0.135	0.088	0.281	0.131	0.119	0.069
$N = 500$, $T = 4$	-0.069	0.009	-0.027	0.005	0.226	0.130	0.098	0.061	0.190	0.103	0.081	0.049
$N = 1000$, $T = 4$	-0.037	0.010	-0.015	0.007	0.170	0.116	0.074	0.052	0.138	0.093	0.059	0.042
$N = 5000$, $T = 4$	-0.007	0.008	-0.003	0.004	0.077	0.069	0.033	0.029	0.061	0.055	0.026	0.024
$N = 100$, $T = 8$	-0.069	0.012	-0.014	0.004	0.081	0.091	0.032	0.037	0.094	0.073	0.029	0.029
$N = 100$, $T = 12$	-0.041	0.003	-0.004	0.001	0.045	0.050	0.020	0.023	0.054	0.039	0.016	0.017
$N = 5000$, $T = 12$	-0.001	0.000	0.000	0.000	0.006	0.005	0.003	0.003	0.005	0.004	0.002	0.002

Notes. AB refers to the Arellano and Bond (1991) GMM estimator; ML refers to the maximum likelihood estimator discussed in Section 2.1; True parameter values are $\lambda = 0.75$ and $\beta = 0.25$; Bias refers to the median estimation errors $\hat{\lambda} - \lambda$ and $\hat{\beta} - \beta$; iqr is the $75th$-$25th$ interquartile range; RMSE is the root mean square error; results are based on 1,000 replications.

Table 2 considers alternative DGPs in which the persistence of the dependent variable is larger than in the baseline design (i.e. λ is closer to 1). Under these circumstances, the AB biases are expected to increase as instruments become weaker (Bond et al., 2001). Indeed, columns (1) and (3) confirm this pattern for both λ and β. In the case of ML in columns (2) and (4), biases are also larger as λ increases but the magnitude of these biases is substantially smaller than that of AB. Turning to efficiency, iqrs tend to increase with λ in columns (5) and (7) for AB, but remain similar or even lower in the case of ML as reported in columns (6) and (8). Finally, RMSEs in columns (9)-(12) summarize these findings pointing to significantly lower RMSEs for the ML estimator. Indeed, the RMSEs of the ML estimator relative to those of the AB estimator are reduced as λ increases: the RMSE for the AB estimator is two times larger than that of ML when $\lambda = 0.75$ and four times larger when $\lambda = 0.99$.

Table 3 explores the performance of our ML estimator when working with unbalanced panels, which are very common in practice. In particular, we consider samples with different degrees of un-

Table 2: Simulation results for different values of λ.

	Bias λ		Bias β		iqr λ		iqr β		RMSE λ		RMSE β	
	AB	ML	AB	ML	AB	ML	AB	ML	AB	ML	AB	ML
Sample size	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\lambda = 0.75$	-0.138	-0.002	-0.054	0.002	0.312	0.167	0.135	0.088	0.281	0.131	0.119	0.069
$\lambda = 0.80$	-0.169	-0.010	-0.067	-0.002	0.339	0.161	0.147	0.087	0.315	0.127	0.133	0.068
$\lambda = 0.85$	-0.208	-0.015	-0.083	-0.002	0.373	0.152	0.162	0.087	0.358	0.120	0.152	0.068
$\lambda = 0.90$	-0.252	-0.029	-0.103	-0.008	0.413	0.150	0.181	0.086	0.409	0.120	0.175	0.068
$\lambda = 0.95$	-0.300	-0.039	-0.125	-0.013	0.455	0.146	0.201	0.086	0.466	0.121	0.200	0.068
$\lambda = 0.99$	-0.335	-0.048	-0.142	-0.018	0.478	0.142	0.211	0.086	0.503	0.121	0.218	0.069
$\lambda = 1.00$	-0.343	-0.052	-0.146	-0.020	0.481	0.142	0.213	0.086	0.509	0.123	0.221	0.070

Notes. AB refers to the Arellano and Bond (1991) GMM estimator; ML refers to the maximum likelihood estimator discussed in Section 2.1; Sample size is $N = 200$ and $T = 4$ in all cases; Bias refers to the median estimation errors $\hat{\lambda} - \lambda$ and $\hat{\beta} - \beta$; iqr is the $75th$-$25th$ interquartile range; RMSE is the root mean square error; results are based on 1,000 replications.

balancedness satisfying the missing at random (MAR) assumption.[7] First, we compute a "probability of missing observation" P_{it}^m that depends on x as follows: $P_{it}^m = \Lambda(0.5x_{it} + \varsigma_{it})$ where $\varsigma_{it} \sim N(0,1)$. Second, both y and x are replaced by missing values for those observations below the $1st$, $5th$ and $10th$ percentiles of the P_{it}^m distribution. Therefore, we explore the performance of the estimators depending on the severity of the unbalancedness.

Two main conclusions emerge from the results in Table 3. First, the larger the severity of the unbalancedness, the larger the finite sample biases. However, in the case of the ML estimator the biases remain much lower in all cases. Second, the $75th$-$25th$ interquartile ranges also increase significantly as the unbalancedness increases. However, the iqr increases are lower in the case of the ML estimator. In any event, we acknowledge that some samples in our simulations produce convergence failures in the ML estimator.[8] All in all, while the ML estimator suffers from convergence

[7] Under MAR, the probability that an observation is missing on variable y can depend on another observed variable x. This condition is thus less restrictive than the missing completely at random (MCAR) assumption that requires missing values on y to be independent of other observed variables x as well as the values of y itself.

[8] The FIML algorithm can fail to converge when working with unbalanced panels, especially with small sample sizes. For example, in Panel A of Table 3 with $N = 200$ and $T = 4$, there was a convergence failure in around 20% of the samples, which were excluded from the results shown in the table. However, this figure is around 10% in Panel B with $N = 500$ and $T = 4$, and less than 1% in Panel C with $N = 200$ and $T = 8$. Indeed, in all samples with $T = 8$ the percentage of failures is less than 1%. Therefore, we conclude that convergence failures of our estimator may be a concern when exploiting unbalanced panels in which time series dimension is low (around $T = 4$) and the share of

problems when unbalancedness is severe and the time dimension is low, the finite sample biases in the AB estimator significantly increase under these circumstances. Note also that Williams et al. (2018) discuss ways to get models to converge when they initially fail to do so.

Table 3: Simulation results under unbalanced panels.

	Bias λ		Bias β		iqr λ		iqr β	
	AB	ML	AB	ML	AB	ML	AB	ML
Unbalacedness	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				PANEL A: $N = 200$, $T = 4$				
1%	-0.171	-0.005	-0.063	0.006	0.336	0.212	0.134	0.099
5%	-0.218	-0.004	-0.082	0.000	0.381	0.212	0.153	0.091
10%	-0.268	0.005	-0.111	0.003	0.381	0.222	0.154	0.100
				PANEL B: $N = 500$, $T = 4$				
1%	-0.090	-0.003	-0.035	-0.003	0.235	0.160	0.100	0.071
5%	-0.122	0.009	-0.051	0.005	0.282	0.155	0.114	0.070
10%	-0.163	0.016	-0.065	0.005	0.307	0.175	0.125	0.074
				PANEL C: $N = 200$, $T = 8$				
1%	-0.049	0.004	-0.009	0.004	0.067	0.067	0.027	0.029
5%	-0.072	0.015	-0.015	0.010	0.081	0.083	0.032	0.034
10%	-0.104	0.020	-0.027	0.014	0.099	0.087	0.042	0.036
				PANEL D: $N = 500$, $T = 8$				
1%	-0.021	0.006	-0.004	0.003	0.043	0.037	0.018	0.017
5%	-0.035	0.014	-0.008	0.007	0.053	0.043	0.021	0.018
10%	-0.054	0.022	-0.015	0.011	0.063	0.048	0.026	0.019

Notes. AB refers to the Arellano and Bond (1991) GMM estimator; ML refers to the maximum likelihood estimator discussed in Section 2.1 implemented based on the FIML approach as described in the main text (i.e. `fiml` option in the `xtdpdml` Stata command); True parameter values are $\lambda = 0.75$ and $\beta = 0.25$; Bias refer to the median estimation errors $\hat{\lambda} - \lambda$ and $\hat{\beta} - \beta$; iqr is the $75th$-$25th$ interquartile range; results are based on 1,000 replications; unbalacedness refers to the share of observations with missing value according to the missing at random (MAR) assumption.

The simulation results discussed in this section are expected to hold under non-normality of the disturbances; this is so because ML can be considered a pseudo maximum likelihood estimator that remains consistent and asymptotically normal under non-normality (see Moral-Benito, 2013). In Table 4, we explore fat-tailed and skew disturbances under different degrees of unbalancedness to check the sensitivity of the FIML-based estimates to the normality assumption, especially in the case

missing values is large (above 10%).

Table 4: Simulation results under nonnormal disturbances.

	Bias λ		Bias β		iqr λ		iqr β	
	AB	ML	AB	ML	AB	ML	AB	ML
Unbalacedness	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				PANEL A: t-student 4 df. $N = 200$, $T = 4$				
0%	-0.165	0.016	-0.063	0.009	0.312	0.190	0.136	0.089
5%	-0.225	-0.002	-0.079	-0.003	0.353	0.199	0.163	0.094
10%	-0.269	-0.017	-0.105	-0.007	0.413	0.189	0.181	0.088
				PANEL B: t-student 4 df. $N = 500$, $T = 4$				
0%	-0.074	0.012	-0.030	0.006	0.237	0.153	0.101	0.066
5%	-0.141	-0.008	-0.056	0.002	0.266	0.136	0.113	0.062
10%	-0.187	-0.012	-0.073	-0.001	0.331	0.154	0.140	0.065
				PANEL C: Mixture of Normals. $N = 200$, $T = 4$				
0%	-0.181	-0.011	-0.080	-0.002	0.335	0.173	0.166	0.088
5%	-0.217	-0.023	-0.054	0.013	0.328	0.166	0.136	0.077
10%	-0.225	-0.030	-0.052	0.005	0.307	0.154	0.115	0.068
				PANEL D: Mixture of Normals. $N = 500$, $T = 4$				
0%	-0.096	0.005	-0.042	0.006	0.247	0.128	0.121	0.063
5%	-0.217	-0.023	-0.054	0.013	0.328	0.166	0.136	0.077
10%	-0.225	-0.030	-0.052	0.005	0.307	0.154	0.115	0.068

Notes. AB refers to the Arellano and Bond (1991) GMM estimator; ML refers to the maximum likelihood estimator discussed in Section 2.1 implemented based on the FIML approach as described in the main text (i.e. `fiml` option in the `xtdpdml` Stata command); True parameter values are $\lambda = 0.75$ and $\beta = 0.25$; Bias refer to the median estimation errors $\hat{\lambda} - \lambda$ and $\hat{\beta} - \beta$; iqr is the $75th$-$25th$ interquartile range; results are based on 1,000 replications; unbalacedness refers to the share of observations with missing value according to the missing at random (MAR) assumption.

of unbalanced panels in which the normality assumption might appear more relevant. In particular, we first consider that all errors in the DGP are distributed as a Student with 4 degrees of freedom (Panels A and B), implying an infinite kurtosis; that is, fatter tails than the normal distribution. Second, we also consider errors distributed according to a mixture of two normal distributions with different means (being the difference equal to 20) so that the resulting distribution is nonsymmetric (Panels C and D). For both nonnormal disturbances, the results remain very similar to those of the normal case.

4 Empirical Application

The growth regressions literature over the eighties and early nineties was mostly based on cross-section approaches (see e.g. Barro, 1991). Starting in the mid-nineties, the mainstream approach was based on panel data methods accounting for country-specific effects and reverse causality between economic growth and potential growth determinants. The Arellano and Bond (1991) estimator was the most popular alternative exploited in this literature (e.g. Caselli et al., 1996).

Along these lines, the influential paper by Levine et al. (2000) found a positive effect of financial development on economic growth after accounting for country-specific fixed effects and reverse causality in a panel data setting. They first considered the Arellano and Bond (1991) first-differenced GMM estimator. However, given the concern of finite sample biases in first-differenced GMM due to the small N dimension of their data (they only observe 74 countries), they also explored the system-GMM approach by Arellano and Bover (1995). The mean stationarity assumption required for consistency of the system-GMM estimator is especially inappropriate in cross-country datasets starting at the end of a war, as argued by Barro and Sala-i-Martin (2003). In this context, the maximum likelihood approach discussed in this paper is a natural alternative to be explored instead of system-GMM.

In this section, we estimate the effect of financial development on economic growth using the proposed ML estimator in addition to first-differenced GMM. We use a panel dataset of 78 countries ($N = 78$) over the period 1960-1995.[9] Following Levine et al. (2000) we consider 5-year periods to avoid business cycle fluctuations so that we exploit a maximum of 7 observations per country ($T = 7$).

The dependent variable is the log of real per capita GDP taken from the World Development Indicators (WDI). The main regressors of interest are three different proxies for financial development at the country level, namely, liquid liabilities, commercial-central bank, and private credit, all taken from the International Financial Statistics (IFS) database. Liquid liabilities are defined as the liquid liabilities of the financial system (currency plus demand and interest-bearing liabilities of banks and non-bank financial intermediaries) divided by GDP. Commercial-central bank is defined as the assets

[9] Since we do not have the original data set assembled by Levine et al. (2000), we use an equivalent data set taken from the same public sources and including four additional countries. We thank Pau Gaya and Alexandro Ruiz for sharing these data with us.

of deposit money banks divided by assets of deposit money banks plus central bank assets. Private credit refers to the credit by deposit money banks and other financial institutions to the private sector divided by GDP. Finally, the following control variables are also considered: opennes to trade (from WDI), government size (from WDI), average years of secondary schooling (from the Barro and Lee dataset), inflation (IFS), and the black market premium (from World Currency Yearbook). For more details on the variables considered see Table 12 in Levine et al. (2000).

Analogously to equations (1)-(2), we estimate the following model:

$$y_{it} = \lambda y_{it-1} + \beta FD_{it} + \gamma w_{it} + \alpha_i + v_{it} \tag{15}$$

where y_{it} refers to the log of real per capita GDP in country i and period t,[10] FD_{it} refers to one of the three financial development proxies considered by Levine et al. (2000), and w_{it} referes to a set of control variables. α_i captures country-specific heterogeneity potentially correlated with the regressors that is time-invariant. In addition, we also include a set of time dummies to account for common shocks to all countries (e.g. the 1973 crisis). β is our parameter of interest, as it estimates the effect of financial development on economic growth.[11]

Following Levine et al. (2000) we assume that both FD_{it} and the control variables w_{it} are predetermined so that feedback from GDP to financial development and other macroeconomic conditions is allowed:

$$E\left(v_{it} \mid y_i^{t-1}, w_i^t, FD_i^t, \alpha_i\right) = 0 \qquad (t = 1, ..., T)(i = 1, ..., N) \tag{16}$$

The Arellano and Bond (1991) approach as well as the likelihood-based approach discussed in this paper can estimate the model in (15) under assumption (16). However, note that the system-GMM estimator, also considered by Levine et al. (2000), requires the additional assumption of mean-stationarity that seems undesirable in this setting as discussed in Barro and Sala-i-Martin (2003).

Table 5 presents the estimation results. In all cases the FIML approach was considered in the ML estimator due to the unbalancedness of the panel.[12] There are 445 observations with non-missing

[10] Note that we consider seven five-year periods, namely, 1960-1965, 1965-1970, 1970-1975, 1975-1980, 1980-1985, 1985-1990, and 1990-1995.

[11] Note that the model in (15) is equivalent to $y_{it} - y_{it-1} = (\lambda - 1)y_{it-1} + \beta FD_{it} + \gamma w_{it} + \alpha_i + v_{it}$ where the dependent variable is GDP growth.

[12] We use the `fiml` option in the `xtdpdml` Stata command.

Table 5: Financial development and economic growth.

PANEL A: First-differenced GMM estimator (AB)						
Lagged dep. variable	0.704***	0.617***	0.731***	0.629***	0.638***	0.579***
	(0.066)	(0.049)	(0.056)	(0.048)	(0.057)	(0.049)
Liquid Liabilities	0.040**	0.066***				
	(0.019)	(0.017)				
Commercial-central bank			0.039***	0.039***		
			(0.011)	(0.010)		
Private Credit					0.050***	0.054***
					(0.013)	(0.015)
Control variables	Simple	Policy	Simple	Policy	Simple	Policy
Observations	417	397	429	398	417	396

PANEL B: Maximum likelihood estimator (ML)						
Lagged dep. variable	1.019***	1.004***	0.980***	0.960***	0.955***	0.945***
	(0.043)	(0.050)	(0.044)	(0.048)	(0.040)	(0.042)
Liquid liabilities	0.029**	0.028**				
	(0.012)	(0.014)				
Commercial-central bank			0.044***	0.041***		
			(0.008)	(0.008)		
Private credit					0.053***	0.048***
					(0.010)	(0.009)
Control variables	Simple	Policy	Simple	Policy	Simple	Policy
Observations	411	397	429	398	417	396

Notes. Dependent variable is the log of real per capita GDP in all cases. Simple set of control variables includes only average years of secondary schooling as an additional covariate. The policy conditioning information set includes average years of secondary schooling, government size, openness to trade, inflation, and black market premium as in Levine et al. (2000). All regressors are normalized to have zero mean and unit standard deviation in order to ease the interpretation of the coefficients. We denote significance at 10%, 5% and 1% with *, ** and ***, respectively. Standard errors are denoted in parentheses.

values in all the variables while the total number of observations is 78×7=546 (i.e. unbalancedness is around 18%).[13] Still, the ML algorithm achieved convergence in all the specifications producing

[13]Note that the inclusion of the lagged dependent variable further reduces the number of observations used in Table 5.

reasonable estimates, which can be attributed to the availability of a relatively large number of time series observations ($T = 7$) as illustrated in our simulation results in Section 3.

The diff-GMM estimates in Panel A of Table 5 replicate the findings in Levine et al. (2000). All the three proxies for financial development (liquid liabilities, commercial-central bank, private credit) have a positive and statistically significant effect on economic growth. Moreover, the effects are economically large since all regressors are normalized to have zero mean and unit standard deviation. For instance, an increase of one standard deviation in the credit-to-GDP ratio boosts the level of GDP per capita by around 5.4% according to the estimates in the last column of Panel A. The magnitude of the liquid liabilities and commercial-central bank estimated effects are also large and similar in magnitude. Also, given the estimated persistence of GDP per capita (i.e. the lagged dependent variable coefficient), the long-run effects are even larger. In particular, the long-run effect on a one standard deviation increase in private credit is estimated to be around 13% (i.e. $\frac{\beta}{1-\lambda}$).

Turning to the maximum likelihood estimates in Panel B of Table 5, the estimated effects are overall very similar. For instance, the estimated impact effect of private credit on GDP per capita is 4.8% instead of 5.4% as in Panel A. However, the estimated coefficients for the lagged dependent variable are significantly larger when using the ML estimator, which points to a downward bias in the diff-GMM estimates as shown in our simulations. An important implication of this result is that the estimated long-run effects of financial development on GDP could be much larger. According to the last column of Panel B, the estimated long-run effect on GDP per capita of a one standard deviation increase in private credit is 87% ($\frac{0.048}{1-0.945}$) instead of 13% ($\frac{0.054}{1-0.579}$) as estimated by diff-GMM. Not surprisingly, all the coefficients are estimated more precisely than in the GMM case as maximum likelihood is more efficient than GMM under normality.

5 Concluding Remarks

The widely-used first-differenced GMM estimator discussed in Arellano and Bond (1991) may suffer from finite sample biases when the number of cross-section observations is small. Based on the same identifying assumption, the alternatives proposed in the literature are typically difficult to implement by practitioners as they require some programming capabilities (e.g. Alonso-Borrego and Arellano (1999), Ahn and Schmidt (1995), Hansen et al. (1996), Hsiao et al. (2002), Moral-Benito

(2013)).[14] Moreover, concern has intensified in recent years that many instrumental variables of the type considered in panel GMM estimators such as Arellano and Bond (1991) may be invalid, weak, or both (see Bazzi and Clemens, 2013; Kraay, 2015).

In this article, we discuss a maximum likelihood estimator that is asymptotically equivalent to the Arellano and Bond (1991) estimator but it is strongly preferred in terms of finite sample performance. Moreover, the proposed estimator can be easily implemented in various SEM packages such as Stata (`xtdpdml` command described in Williams et al. (2018)), SAS (proc CALIS), Mplus, LISREL, EQS, Amos, lavaan (for R), and OpenMx (for R).

Simulation results presented in the paper indicate that our maximum likelihood estimator has negligible biases in finite samples when the DGP includes fixed effects, a lagged dependent variable regressor, and an additional predetermined explanatory variable. Moreover, these biases are smaller than those of first-differenced GMM when the number of cross-section observations (N) is small.

As an empirical illustration, we estimate the effect of financial development on economic growth in a panel of countries using the proposed estimator. According to our empirical results, the GMM estimates of the long-run effect of financial development on economic growth presented in Levine et al. (2000) are much larger when considering the proposed maximum likelihood estimator.

References

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

[2] Allison, P. (2005) "Fixed Effects Regression Methods for Longitudinal Data Using SAS," SAS Institute Inc, Cary, NC.

[3] Allison, P., R. Williams, and E. Moral-Benito (2017) "Maximum Likelihood for Cross-Lagged Panel Models with Fixed Effects," Socius, 3, 1-17.

[4] Alonso-Borrego, C. and M. Arellano (1999) "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data," Journal of Business & Economic Statistics, 17, 36-49.

[14] Note that the so-called system-GMM estimator by Arellano and Bover (1995) and Blundell and Bond (1998) is not included in this category because it requires the mean-stationarity assumption for consistency, which is not required by first-differenced GMM.

[5] Arbuckle, J. (1996) “Full information estimation in the presence of incomplete data,” In G. A. Marcoulides & R. E. Schumacker (Eds.), Advanced structural equation modeling (pp. 243277). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

[6] Arellano, M. (2003) “Panel Data Econometrics,” Oxford, UK: Oxford University Press

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

[8] Arellano, M. and O. Bover (1995) “Another Look at the Instrumental Variable Estimation of Error-Components Models,” Journal of Econometrics, 68, 29-52.

[9] Barro, R. (1991) “Economic Growth in a Cross Section of Countries,” Quarterly Journal of Economics, 106, 407-443.

[10] Barro, R. and X. Sala-i-Martin (2003) “Economic Growth.” MIT Press: Cambridge, MA.

[11] Bazzi, S. and M. Clemens (2013) “Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth,” American Economic Journal: Macroeconomics, 5, 152186.

[12] Bentler, P. and D. Weeks (1980) “Linear structural equations with latent variables,” Psychometrika, 45, 289-308.

[13] Blundell, R. and S. Bond (1998) “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models,” Journal of Econometrics, 87, 115-143.

[14] Bond, S., A. Hoeffler, and J. Temple (2001) “GMM Estimation of Empirical Growth Models,” CEPR Discussion Papers 3048.

[15] Bun, M. and J. Kiviet (2006) “The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models,” Journal of Econometrics, 132, 409-444.

[16] Caselli, F., G. Esquivel and F. Lefort (1996) “Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics,” Journal of Economic Growth, 1, 363-389.

[17] Enders, C. and D. Bandalos (2001) "The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models," Structural Equation Modeling, 8, 430-457.

[18] Hansen, L. P., J. Heaton, and A. Yaron (1996) "Finite-Sample Properties of Some Alternative GMM Estimators," Journal of Business & Economic Statistics, 14, 262-280.

[19] Hsiao, C., H. Pesaran, and A. Tahmiscioglu (2002) "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," Journal of Econometrics, 109, 107-150.

[20] Islam, N. (1995) "Growth Empirics: A Panel Data Approach," The Quarterly Journal of Economics, 110, 1127-1170.

[21] Kraay, A. (2015) "Weak Instruments in Growth Regressions Implications for Recent Cross-Country Evidence on Inequality and Growth," Policy Research Working Paper 7494.

[22] Levine, R., N. Loayza, and T. Beck (2000) "Financial Intermediation and Growth: Causality and Causes," Journal of Monetary Economics, 46, 31-77.

[23] Moral-Benito, E. (2013) "Likelihood-based Estimation of Dynamic Panels with Predetermined Regressors," Journal of Business & Economic Statistics, 31, 451-472.

[24] Williams, R., P. Allison, and E. Moral-Benito (2018) "xtdpdml: Linear Dynamic Panel-Data Estimation using Maximum Likelihood and Structural Equation Modeling," The Stata Journal, forthcoming.

[25] Wooldridge, J. (2010) "Econometric Analysis of Cross Section and Panel Data," The MIT Press. Cambridge, MA.

A Illustration in the Case of Three Time Periods

In order to illustrate the equivalence of our likelihood-based approach outlined in section 2.1 and the baseline GMM approach exclusively based on assumption (2), we consider the case $T = 3$ and show that the number of over-identifying restrictions is the same in both estimators.

A.1 The GMM approach

With three time periods and y_{i0} observed by the econometrician, the model in (1)-(2) implies the following moment conditions:

$$E(y_{i0}\Delta v_{i2}) = 0 \tag{17a}$$
$$E(x_{i1}\Delta v_{i2}) = 0 \tag{17b}$$
$$E(y_{i0}\Delta v_{i3}) = 0 \tag{17c}$$
$$E(y_{i1}\Delta v_{i3}) = 0 \tag{17d}$$
$$E(x_{i1}\Delta v_{i3}) = 0 \tag{17e}$$
$$E(x_{i2}\Delta v_{i3}) = 0 \tag{17f}$$
$$E(\Delta v_{i2}(v_{i3} + \alpha_i)) = 0 \tag{17g}$$

The moments (17a)-(17b) are those typically exploited by first differenced GMM as in Arellano and Bond (1991) while the moment in (17g) results from the lack of autocorrelation implied by assumption (2) as considered by Ahn and Schmidt (1995).

We thus have seven moment conditions and two parameters to be estimated, λ and β, which give rise to five over-identifying restrictions implied by the model in (1)-(2) when λ and β are the parameters of interest.

A.2 The likelihood-based approach

The model in structural form given by equation (10) involves 23 structural parameters when $T = 3$, namely, λ, β, σ^2_α, σ^2_{v0}, σ^2_{v1}, σ^2_{v2}, σ^2_{v3}, $\sigma^2_{\xi_1}$, $\sigma^2_{\xi_2}$, $\sigma^2_{\xi_3}$, ϕ_0, ϕ_1, ϕ_2, ϕ_3, ψ_{21}, ψ_{31}, ψ_{32}, ω_{01}, ω_{02}, ω_{03}, ω_{12}, ω_{13} and ω_{23}.

The reduced form version of the model in (10), given by $R_i = B^{-1}DU_i$, involves 28 reduced form parameters coming from the 7×7 covariance matrix of the reduced-form disturbances $\Xi_i = B^{-1}DU_i$.

The difference between 28 reduced form parameters and 23 structural parameters implies 5 over-identifying restrictions as in the GMM case above, which ensures identification and that our likelihood-based approach does not impose any additional restriction (i.e. it is exclusively based on assumption (2)).