Political Science Research and Methods (2020), page 1 of 20
doi:10.1017/psrm.2020.30

ORIGINAL ARTICLE

Transformed-likelihood estimators for dynamic panel models with a very small *T*

Mark Pickup and Vincent Hopkins

Department of Political Science, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
Corresponding author. Email: mark.pickup@sfu.ca

(Received 2 December 2018; revised 27 June 2019; accepted 8 October 2019)

Conventional OLS fixed-effects and GLS random-effects estimators of dynamic models that control for individual-effects are known to be biased when applied to short panel data ($T \leq 10$). GMM estimators are the most used alternative but are known to have drawbacks. Transformed-likelihood estimators are unused in political science. Of these, *orthogonal reparameterization* estimators are only tangentially referred to in any discipline. We introduce these estimators and test their performance, demonstrating that the unused orthogonal reparameterization estimator in particular performs very well and is an improvement on the commonly used GMM estimators. When T and/or N are small, it provides efficiency gains and overcomes the issues GMM estimators encounter in the estimation of long-run effects when the coefficient on the lagged dependent variable is close to one.

Keywords: Bayesian; dynamic; individual effects; panel data; transformed-likelihood; unobserved heterogeneity

Panel data allow researchers to control for unobservable individual-effects and effects of the past that cross-sectional data can neither identify nor control (Finkel, 2008; Hsaio, 2014, 4–10). But researchers must attend to the model and estimation technique used (Stimson 1985). Researchers must account for the dynamic structure of their data and think carefully about controlling for individual-effects (Beck and Katz, 1996, 2001; Green *et al.*, 2001; Wawro, 2002; Wilson and Butler, 2007). Past literature reviews show that researchers have done a poor job of this (Wawro, 2002; Wilson and Butler, 2007). Our own more recent literature review suggests things have improved.[1] Researchers using panel data are more likely to address issues of autoregressive data or individual-effects but rarely both. This is a concern because solutions to one issue that do not consider the other can actually make point estimates worse (Plümper and Troeger, 2019).

When researchers do address both autoregressive data and individual-effects, there is a lack of consensus on how to do so. This, in part, is because the otherwise very helpful political methodology literature on how to deal with unobserved individual-effects (inter alia, Shor et al. 2007; Zhu, 2012; Bell and Jones, 2015; Clark and Linzer 2015) within a (autoregressive) dynamic model (inter alia, Beck and Katz, 1996, 2001; Green *et al.*, 2001; Wawro, 2002; Wilson and Butler, 2007) provides insufficient guidance on this for short panel data—that is, data with as few as three waves and no more than ten. This is typical for but not limited to public opinion and political behavior data[2] and 57 percent of the articles in our literature review used panel

[1]This is based on a review of articles using panel data in the American Political Science Review, American Journal of Political Science, Journal of Politics, Public Opinion Quarterly and Political Behavior (January 2010 to June 2018). The list of articles is available in Appendix A.

[2]This is the structure of most national election panel studies, social surveys and socioeconomic panels (e.g., British Election Panel Study 2016; American National Election Study 2000-2002-2004 Panel).

data in this range.[3] Under these circumstances, conventional estimators such as ordinary least squares fixed-effects (OLS-FE) and standard generalized least squares random-effects (GLS-RE) are biased (Nickell, 1981; Hsiao, 2014), and common alternative generalized methods of moments (GMM) estimators can produce large random errors when T and/or N are small, and are sometimes biased (Blundell and Bond, 1998, Bond and Windmeijer, 2002). This is particularly true for the estimation of the long-run effect (LRE). Meanwhile, much less well-known transformed-likelihood estimators (TLEs) for dynamic panel models have recently become available. This can leave a researcher uncertain how to proceed.

This paper makes three contributions. First, it introduces TLEs for very small T, dynamic, fixed-effects models designed to address the problems of OLS-FE, GLS-RE, and GMM estimators. TLEs are unknown and hitherto unused in political science. Of these, the *orthogonal reparameterization* estimator (OPM) is also virtually unknown outside of political science. This is likely because it is only discussed in the most abstract terms—as an example demonstrating an approach to the statistical problem of incidental parameters (Lancaster, 2002). It has been argued that the better-known TLE, the *quasi-maximum likelihood* estimator (QML-FE), is an improvement over GMM estimators in that it has better finite sample properties but that it runs into difficulties estimating the LRE of model covariates when T and N are small and the coefficient on the lagged dependent variable (LDV) is greater than 0.8 (Hsiao *et al.*, 2002).[4] The performance of OPM is unknown when the number of cases is not large—common in political science.[5] The second contribution of this paper then is to test the finite sample performance of TLEs. We demonstrate TLEs provide efficiency gains over GMM, and OPM in particular has very good finite sample properties and is an improvement over both QML-FE and GMM estimators, particularly when estimating the LRE. The third contribution is to describe how the gains made by TLEs are achieved by making distributional assumptions and to test the performance of these estimators when these distributional assumptions are badly violated. The robustness of TLEs to such violations is so far unknown. We demonstrate that OPM in particular is largely robust to common violations. Overall, OPM is always as good as the estimators currently being used and in circumstances important to political science researchers it is often better.

In making these contributions, this paper focuses on linear, single equation models estimated on short panels ($T \leq 10$) with correlations within individuals but not across them.[6] We also assume that our hypotheses are focused on the effects of changes in time-varying independent variables. Appendix B provides a discussion of these assumptions.[7] We make additional assumptions about the homogeneity of effects over-time and across cases that may or may not hold with a particular dataset. Violations of over-time homogeneity have little effect when T is very small, as it does not have an opportunity to present itself. The potential for violations of cross-case homogeneity is a limitation that is discussed in the concluding section.

1 Unobservable individual-effects

The simplest panel data model is the pooled model:

$$y_{i,t} = \beta_0 + \beta_1 x_{i,t} + \mu_{i,t} \tag{1}$$

[3]A further 16 percent had only two waves. And the remaining had more than ten waves.

[4]T equals the number of waves of data, including the first wave which we lose in order to have a measure of the LDV in the second wave.

[5]In about 38 percent of the articles in our literature review with ten or fewer waves, N was less than 500 and in almost half of these, N was less than 50.

[6]Or at least not once we control for time-specific effects.

[7]We have not made any assumptions regarding the stationarity of the data. This is because in political science data, once nonstationary elements such as trending, structural breaks, and periodicity are controlled for in the model, dependent variables of interest often exhibit stationarity. Further, if T is very small, the bias due to $I(1)$ nonstationarity is negligible.

where the subscript i indexes the case/individual and subscript t indexes time, so that $y_{i,t}$ is the value of y for case i at time t.

The $\mu_{i,t}$ represent the errors in the model. Problems of unobservable individual-effects occur when the data generating process (DGP) for the errors is as follows:[8]

$$\mu_{i,t} = \eta_i + \epsilon_{i,t} \tag{2}$$

In Equation 2, the structure of the errors is such that they contain a component that differs from case to case but does not vary over time: η_i. These unobservable individual-effects (also known as unobserved heterogeneity) are time-invariant differences between cases not reflected in $x_{i,t}$.[9] These unexplained differences may be characteristics of the cases that do not vary over time or that do vary over time but have different case specific averages over the period of the study. Such unobserved differences are plausible in most research situations (Rosenbaum, 2005; Wilson and Butler, 2007).[10]

The individual-effects (η_i) have two consequences. First, they result in correlation in the errors across time leading to incorrect estimates for the standard errors of the parameters (e.g., the slopes) in the model. The second consequence of the individual-effects occurs if the DGP for $x_{i,t}$ also contains time-invariant differences: $x_{i,t} = \eta_i^x + \nu_{i,t}$

If $Corr(\eta_i \eta_i^x) \neq 0$, the individual-effects in $\mu_{i,t}$ are not independent of $x_{i,t}$. Consequently, the zero conditional mean (exogeneity) assumption, $E(\mu_{i,t}|x_{i,t}) = 0$, is violated and a pooled-OLS estimate of Equation 1 is biased. With individual-level data, this is often expected to be the case (Finkel, 2008).

The appropriate model to resolve the problems of unobservable individual-effects depends on the assumptions we can make about the relationship between the individual-effects in $\mu_{i,t}$ and the covariates in the model—specifically, whether the individual-effects are independent of the covariates (Hausman and Taylor, 1981): $E(\eta_i\, x_{i,t}) = 0$. If we can make this assumption, the η_i are commonly called random-effects and typically a GLS-RE estimator (Balestra and Nerlove, 1966) is used. Otherwise, the η_i are commonly called fixed-effects and typically an OLS-FE estimator is used.

2 The need for dynamic panel models

A dynamic process is one that includes one or more lags of the dependent variable on the right-hand side. For example:

$$y_{i,t} = \alpha y_{i,t-1} + \beta_1 x_{i,t} + \mu_{i,t} \tag{3}$$

$$\mu_{i,t} = \eta_i + \epsilon_{i,t} \quad \epsilon_{i,t} \sim \text{NID}(0,\, \sigma_\epsilon^2)$$

The DGP described by Equation 3 contains a single lag of $y_{i,t}$, reflecting the fact that $y_{i,t}$ is autoregressive. Higher order lags of $y_{i,t}$ are also possible but are very rarely employed in political science. For ease of exposition, we will assume only the first lag is of concern. A dynamic model may also include one or more lags of the independent variables.

A dynamic model is necessary if $y_{i,t}$ is autoregressive. We expect autoregression if we believe that subsequent to something happening that temporarily changes the value of the dependent

[8]Throughout we will assume: $\epsilon_{i,t} \sim \text{NID}(0,\, \sigma_\epsilon^2)$; $E(\epsilon_{i,t}|x_{i,}, z_{m,i,t}) = 0$; $E(\eta_i \epsilon_{i,t}) = 0$; and $E(\epsilon_i \epsilon_j) = \begin{cases} \sigma_\epsilon^2 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$.

[9]When we say these differences are invariant, we only mean for the period of data collection.

[10]The error structure may also contain time-specific effects that affect all the individual cases in more or less the same way. For ease of exposition, we relegate consideration of time-specific effects to the discussion of the "no correlation across cases" assumption in Appendix B.

variable (e.g., a temporary shift in an independent variable), the dependent variable will return partly but not entirely to its original value before the next observation. This autoregressive process is due to shocks in covariates (observed or unobserved) not dissipating by the next observation. For example, if we measure support for the president monthly, we might expect an event (such as a policy announcement) to produce a temporary shift in support. That announcement may still be in the minds of individuals the month after it was made and so the shift in support observed in the month of the announcement may not have entirely dissipated by the next monthly observation. There are often strong substantive reasons to expect such dynamics (Beck and Katz, 1996; Finkel, 2008). In public opinion data in particular, the degree of autoregression is expected to be high (i.e., $\alpha \to 1$ in Equation 3).There may also be reasons to expect lags of some or all of the independent variables in the DGP, reflecting delayed effects on the dependent variable.

To evaluate the consequences of not accounting for an autoregressive dependent variable, consider the estimation of the following static data model:

$$y_{i,t} = \beta_1 x_{i,t} + \eta_i + \varepsilon_{i,t} \tag{4}$$

when the true DGP is Equation 3.

An unbiased OLS-FE or GLS-RE estimation when T is small assumes: $E(\varepsilon_{i,t_1}|x_{i,t_2}) = 0$, for all t_1 and t_2. The expected value of the error is assumed to be independent of any past, current, or future value of the covariates included in the model—strict exogeneity. A problem arises because Equation 4 does not include $\alpha y_{i,t-1}$ and so, the error term $\varepsilon_{i,t}$ is generated by: $\varepsilon_{i,t} = \alpha y_{i,t-1} + \epsilon_{i,t}$

By Equation 3, $y_{i,t-1}$ and $x_{i,t-1}$ are not independent (unless $\beta_1 = 0$) and since $\varepsilon_{i,t}$ is a function of $y_{i,t-1}$, $\varepsilon_{i,t}$ and $x_{i,t-1}$ are not independent. Therefore: $E(\varepsilon_{i,t_1}|x_{i,t_2}) \neq 0$ for $t_2 = t_1 - 1$ and the estimate of β_1 is biased.

A secondary problem is that the errors will be serially correlated: $E(\varepsilon_{i,t}|\varepsilon_{i,t-1}) \neq 0$. Upon noting the serial correlation in the errors, it might be tempting to use a feasible generalized least squares estimator that accounts for the serial correlation (e.g. Prais–Winsten). Unfortunately, this does not resolve the problem of bias produced by not including $y_{i,t-1}$, in the model.[11]

3 The problem of unobservable individual effects in dynamic panel data

It is not the purpose of this paper to suggest that the inclusion of a LDV is always appropriate. However, given the problems created by not using a dynamic model when the DGP requires it, we need such models in our toolbox. Unfortunately, the problems of unobservable individual-effects are magnified by dynamic models.

In static models, if we can assume that the individual-effects are independent of the covariates, $E(\eta_i\ x_{i,t}) = 0$, the primary problem with ignoring the individual-effects is serial correlation in the errors. In a dynamic model, it is not possible to make this assumption because individual-effects cannot be independent of the LDV by definition. From Equation 3, $y_{i,t-1} = \alpha y_{i,t-2} + \beta_1\ x_{i,t-1} + \eta_i + \epsilon_{i,t-1}$. The LDV is a function of the individual effects. Consequently, the OLS estimation of the coefficient on the LDV is biased upwards when individual effects are not included in the data model—i.e. the pooled model (Nerlove, 1967; Trognon, 1978). This has consequences for the estimates of the other coefficients and the LRE. As this independence assumption is also necessary for GLS-RE (Balestra and Nerlove, 1966), its violation means this estimator is also biased in a dynamic model (Hsiao, 2014). Specifically, it is biased toward the pooled model estimates. If the

[11]Note that a model with a LDV will result in a biased OLS estimation when the errors are serially correlated. Wooldridge (2010, 194–196) argues that the solution is to add additional lags of the dependent or independent variables in order to eliminate the serial correlation.

violation of the independence assumption is small, as may be the case in a static model, the bias of the GLS-RE estimator may also be small (Clarke and Linzer, 2015). However, because the assumption of independence is badly violated by the LDV, the bias of traditional GLS-RE for dynamic models can be substantial when T is very small—as we will see.[12]

It is also well-established that under these circumstances, OLS-FE suffers from bias (Nerlove, 1967). When T is small, events that affect $y_{i,t}$ just prior to the first observation will not have fully dissipated before the last observation. This will appear similar to fixed differences between observations, making it hard to distinguish between the fixed-effects and the autoregressive nature of the data. It has been shown that if the true autoregressive parameter, α, is positive, the bias will be negative even as α goes to zero (Nickell, 1981).

The political science literature contains useful advice on how to address individual-effects with dynamic models. Green *et al.* (2001) make a case for the need to control for individual-effects and they include OLS-FE dynamic models amongst the different panel models they estimate. However, they do not include a discussion of the bias that occurs when T is very small, except to note that they only include cases with a T greater than 20 to avoid the problem. Responding to Green *et al.* (2001), Beck and Katz (2001) recommend a random coefficients model as an alternative (Beck and Katz, 2001, 493). In terms of dealing with individual-effects, this is equivalent to a random-effects estimation. For the moderately large T panel data considered in these two papers, this may be good advice but, again, the traditional GLS-RE estimator applied to a dynamic model exhibits substantial bias when $T \leq 10$. Elsewhere, Beck and Katz (2011) suggest that when estimating a dynamic model, the bias from an OLS-FE estimator is relatively small when $T > 20$, but do not discuss the substantial bias when $T \leq 10$. Zhu (2012) usefully notes the problem of estimating dynamic models with a small T but her empirical example has a T of 17 avoiding (as she points out) the problems of a very small T. Wilson and Butler (2007, 107–108) outline very well the problems of estimating a dynamic model with fixed-effects with a very small T. Unfortunately, they do not provide a solution to the problem. They note that the bias of OLS-FE is much larger for the coefficient on the LDV than on the other covariates in the model. As we will show below, the error on the coefficient for all covariates can be substantial when T is very small (≤ 10). Further, bias in estimating the LRE is not a trivial matter. We are typically interested in all effects of $x_{i,t}$ on $y_{i,t}$, not just the short-run effect (SRE). The SRE is often small in magnitude, making it hard to detect, and the LRE is typically larger than the SRE. A downward bias on the estimate of α and so a downward bias on the LRE (see Appendix C) makes it hard to detect the LRE. Overall, this increases type II errors. Finkel (2008)—in brief—outlines the family of GMM estimators proposed by Arellano and Bond (1991), Arellano and Bover (1995) and Blundell *et al.* (2000). These GMM estimators are designed to account for individual-effects in dynamic models without the problems of OLS-FE and GLS-RE when T is small. Wawro (2002) provides an excellent description of these GMM estimators, spelling out the importance of testing the assumptions of the estimator. However, the GMM approach is not without problems, as we discuss below.

This is what the political science literature has covered. What it has not explored, to date, is the potential use of likelihood-based estimators for dynamic models that account for individual-effects. The challenge is that a consistent maximum likelihood estimation (MLE) relies on asymptotics as the number of observations relative to the number of parameters goes to infinity (i.e., a large number of observations relative to the number of parameters). When T is small, the only potentially available asymptotics are in N. However, in a fixed-effects model, for every

[12]Bell and Jones (2015) point out that Mundlak (1978), with certain assumptions about the relationship between the individual-effects and the covariates, unifies the random and fixed-effects approaches for static models. However, this unification does not apply to dynamic models (Hsiao, 2014, 50–52).

case added, a parameter (the individual effect) is also added.[13] Therefore, there are no asymptotics available in N. As we are not interested in the values of the fixed-effects, we call them "incidental parameters" and the challenge they pose for a consistent MLE an "incidental parameters problem" (Neyman and Scott, 1948; Lancaster, 2000).

We examine two likelihood-based estimators as a solution. These are the quasi-maximum likelihood fixed effects (QML-FE) estimator (Hsiao *et al.*, 2002) and the orthogonal reparameterization likelihood (OPM) estimator (Lancaster, 2002). In both cases, the likelihood is transformed so that the incidental parameters are no longer included within it.

Even though these TLEs were proposed in 2002, they are far less known than GMM estimators. The papers introducing QML-FE (Hsiao *et al.*, 2002) and OPM (Lancaster, 2002) have only 81 and 50 cites respectively (Web of Science), and citations of Lancaster are not for applications of OPM but rather reviews of the incidental parameters problem. The two papers that introduced the GMM estimators (Arellano and Bond, 1991; Arellano and Bover, 1995) have 9069 cites between them. This is in some part because statistical packages for TLEs have not been available. To help address this specific issue, we have made OPM available in R (Pickup *et al.*, 2017).[14] The lack of familiarity with OPM is also because the estimator is only discussed as a proof of concept for a general approach to solving the incidental parameters problem (Lancaster, 2002), its performance is only given brief consideration and there is a typographical error in the equation describing the estimator.[15] The general performance of OPM is not well understood. An additional barrier to the adoption of TLEs is that the finite sample properties (small N and T) of OPM and the robustness of both TLEs to violations of distributional assumptions are unknown. We address these issues now by introducing the reader to TLEs for very small T dynamic panel models, testing their performance against the more common GMM estimator, and explaining their trade-offs.

4 Transformed-likelihood estimators for very small *T* dynamic panel models

Before introducing TLEs, we briefly describe GMM estimators. Consider the following first difference transformation. If we assume the DGP in the process described by Equation 3, it is also true that:

$$y_{i,t-1} = \alpha y_{i,t-2} + \beta x_{i,t-1} + \eta_i + \epsilon_{i,t-1} \tag{5}$$

If we subtract Equation 5 from Equation 3 we end up with an equation in first differences and without the individual effects:

$$y_{i,t} - y_{i,t-1} = \alpha y_{i,t-1} - y_{i,t-2} + \beta(x_{i,t} - x_{i,t-1}) + \eta_i - \eta_i + \epsilon_{i,t} - \epsilon_{i,t-1}$$

or

$$\Delta y_{i,t} = \alpha \Delta y_{i,t-1} + \beta \Delta x_{i,t} + \Delta\epsilon_{i,t} \tag{6}$$

Note: $\epsilon_{i,t-1} = y_{i,t-1} - \alpha y_{i,t-2} - \beta x_{i,t-1} - \eta_i$ by Equation 5. As both $\Delta\epsilon_{i,t}$ and $\Delta y_{i,t-1}$ are functions of $y_{i,t-1}$, $E(\Delta\epsilon_{i,t}|\Delta y_{i,t-1}) \neq 0$, which is a requisite for an unbiased estimate of Equation 6. Anderson and Hsiao (1982) propose using the second lag of either the first difference or level of the dependent variable ($\Delta y_{i,t-2}$ or $y_{i,t-2}$) as an instrument for the first differenced LDV, $\Delta y_{i,t-1}$, in Equation 6. Arellano and Bond (1991) proposed a generalized method of moments estimator based on a similar principle. To understand the GMM approach, consider a simple linear regression: $y_{i,t} = \beta x_{i,t} + \epsilon_{i,t}$

[13]The FE estimator is often conducted by first mean differencing the data. Here, it is the need to estimate an additional mean for each additional case that creates the incidental parameters problem.

[14]QML-FE is now available through xtdpdqml in Stata (Kripfganz, 2016).

[15]The stated posterior does not include the priors.

If we believe we can assume that $x_{i,t}$ is independent of $\epsilon_{i,t}$, this suggests the following population moment condition:[16] $E(x_{i,t}\epsilon_{i,t}) = E(x_{i,t}(y_{i,t} - \beta x_{i,t})) = 0$

The GMM estimation of β is calculated by replacing the assumed (and unobservable) population moment condition with the observable sample moment condition:

$$\frac{1}{(N \times T)} \sum_{i,t} x_{i,t}(y_{i,t} - \hat{\beta} x_{i,t}) = 0$$

and solving for $\hat{\beta}$.

If we believe that $x_{i,t}$ and $\epsilon_{i,t}$ are correlated, but that we have an instrument $z_{i,t}$, we can define the following population moment condition: $E(z_{i,t}\epsilon_{i,t}) = E(z_{i,t}(y_{i,t} - \beta x_{i,t})) = 0$ and use this to produce the GMM estimation of β.

These two examples produce estimators that are no different than OLS and instrumental variables regression. However, when more moment conditions are defined, there may be more moment conditions than unknown parameters. GMM proceeds by selecting the values for the parameters in the model that minimizes a weighted sum of the squared moment conditions. The intuition is to select values of the parameters that come closest to meeting the moment conditions on average. This requires estimating a weighting matrix and the efficiency (the relative variance of the estimator) of GMM depends, in part, on the matrix selected.[17]

Arellano and Bond style GMM estimators for Equation 3 use moment conditions based on the levels of lagged variables as instruments in the first differenced Equation 6. The number of available moment conditions is determined by the number of waves of data. The moment conditions used are based on the assumptions the researcher is willing to make about the underlying DGP. The advantage of the Arellano and Bond GMM estimator is that it allows for a greater number of moment conditions than the Anderson and Hsiao approach, reducing random errors.

GMM is susceptible to the equivalent of the weak instruments problem (Stock *et al.*, 2002). When the instrument correlates weakly with the instrumented variable, the estimator is biased toward the OLS estimate of Equation 6 and the sampling distribution is non-normal resulting in incorrectly estimated standard errors (and confidence intervals). The instruments in levels used by Arellano and Bond (1991) are weak when the autoregressive parameter α is too close to 0 or 1 (Blundell and Bond, 1998) or the difference between cases is large relative to changes within cases (between variance is large relative to within variance) (Blundell and Bond, 1998). An extension to the Arellano and Bond GMM method was developed by noting that more moment conditions can be made available (Arellano and Bover, 1995; Blundell and Bond, 1998; Blundell *et al.*, 2000) by using a system GMM estimator. This estimator defines moment conditions based on level instruments for the equation in first differences (Equation 6) and first difference instruments for the original equation in levels (Equation 3). In addition to making more moment conditions available, the first difference instruments can be stronger than the level instruments when α is approximately 1.

Another issue with GMM is that the optimal weighting matrix can be difficult to estimate with limited information (Roodman, 2009), particularly for the system GMM estimator. This occurs when moments are based on weak instruments or when the number of moment conditions is large relative to N. The computational difficulty in estimating the matrix can produce (somewhat unpredictable) bias (Doran and Schmidt, 2006) and reduce efficiency (increasing errors). This means larger standard errors and a greater probability of large random errors. Further, the many instruments problem on its own can result in bias. This occurs when the instrumented variable is over fitted as a consequence of the large number of instruments relative to N. As a result, the instrumenting purges less of the endogenous part of the instrumented variable and

[16] An assumption regarding the joint distribution of variables within the DGP.

[17] For a discussion of the optimal weighting matrix, see Pesaran (2015) section 10.5.

the estimation is again biased toward the OLS estimate. The result is that GMM has poor small N (finite sample) properties (Bond and Windmeijer, 2002). A final limitation of GMM is that the researcher can make many different choices regarding the moment conditions and these choices can produce substantially different results. This opens up the possibility of researchers selecting moment conditions that produce the results that match their theoretical expectations (a researcher degrees of freedom problem). Overall, GMM estimators are an improvement over OLS-FE and GLS-RE but they have limitations.

4.1 The OPM approach

As has been noted, the MLE of $y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \eta_i + \epsilon_{i,t}$ when T is small leads to an incidental parameters problem. Lancaster (2002) suggests a solution. We are not interested in the incidental parameters η_i. We are interested in estimates of the common parameters: α and β (used to calculate the LREs and SREs of $x_{i,t}$) and σ^2. We seek a reparameterization of the incidental parameters so that the incidental and common parameters are information orthogonal. To understand what reparameterization means, let us start with a simple example. If we have the following static model:

$$y_{i,t} = \beta_0 + \beta_1 x_{i,t} + \eta_i + \epsilon_{i,t} \tag{7}$$

we can define a new set of incidental parameters as a function of the original incidental parameters and β_0: $\tilde{\eta}_i = \eta_i + \beta_0$

Our static model, expressed in terms of the new incidental parameters is:

$$y_{i,t} = \tilde{\eta}_i + \beta_1 x_{i,t} + \epsilon_{i,t}$$

Note that although the incidental parameters have changed and their interpretation differs, the common parameters have not changed and their interpretation remains the same.

The orthogonal reparameterization proposed by Lancaster (2002) is much more complex but like the above transformation, it changes the meaning of the parameters representing the individual effects but not the meaning of the other parameters. Unlike the above, the Lancaster reparameterization also allows us to write a likelihood in which the incidental parameters are *informationally orthogonal* from the other parameters. To break this concept down, we begin with the concept of orthogonality.

For Equation 3, we denote the likelihood function for the data for a single case as $\ell_i(\eta_i, \alpha, \beta, \sigma^2)$. Following Lancaster (2002), suppose the incidental parameters can be reparameterized so that the likelihood function factors as:

$$\ell_i(\eta_i, \alpha, \beta, \sigma^2) = \ell_{i1}(\eta_i)\ell_{i2}(\alpha, \beta, \sigma^2) \tag{8}$$

where ℓ_{i1} and ℓ_{i2} are themselves likelihood functions. If the parameters (η_i) and $(\alpha, \beta, \sigma^2)$ are also variation independent, they are orthogonal.[18] It can then be shown that the application of maximum likelihood to the product of the ℓ_{i2} for all cases produces consistent estimates of $(\alpha, \beta, \sigma^2)$ as $N \to \infty$ when $T \geq 3$ (Lancaster, 2002). Lancaster (2002) provides the Poisson count model as an example of where this is possible. The likelihood of case i is:

$$\ell_i(\eta_i, \theta) \propto \mathrm{e}^{-\eta_i \sum_t \exp(x_{i,t}\theta)} \eta_i^{\sum_t y_{i,t}} \mathrm{e}^{\theta_i' \sum_t y_{i,t} x_{i,t}} \tag{9}$$

[18]Informally, variation independence means that in the multidimensional sampling distribution of the model parameters, the covariance between η_i and the other parameters is zero.

If we define $\tilde{\eta}_i = \eta_i \sum_t e^{\theta_i' x_{i,t}}$, we can rewrite Equation 10 as:

$$\ell_i(\tilde{\eta}_i, \theta) \propto e^{-\tilde{\eta}_i} \tilde{\eta}_i^{\sum_t y_{i,t}} \times \frac{e^{\theta_i' \sum_t y_{i,t} x_{i,t}}}{\left(\sum_t e^{\theta_i' x_{i,t}}\right)^{\sum_t y_{i,t}}} \tag{10}$$

This is an orthogonal reparameterization with respect to the incidental parameters because it expresses the original likelihood as the product of two likelihoods, the first of which, $\ell_{i1} = e^{-\tilde{\eta}_i} \tilde{\eta}_i^{\sum_t y_{i,t}}$, contains only the reparameterized incidental parameters and the second, $\ell_{i2} = e^{\theta_i' \sum_t y_{i,t} x_{i,t}} / (e^{\theta_i' \sum_t x_{i,t}})^{\sum_t y_{i,t}}$, only contains the common parameter θ Lancaster (2002).

Unfortunately, not all likelihoods can be transformed so that the incidental parameters are completely orthogonal. However, it may still be possible to reparameterize the incidental parameters so that they are *information* orthogonal. Informally, information orthogonality means no information about the values of the $\tilde{\eta}_i$ tell us anything about the values of the common parameters. Formally, if the incidental parameters are completely orthogonal as above, the following is true:

$$\frac{\partial^2 \mathcal{L}_i}{\partial \tilde{\eta}_i \partial \theta} = 0 \tag{11}$$

where $\mathcal{L}_i$ is the log of the likelihood for the data for case i. Information orthogonality is a weaker form of orthogonality where the above is true in expectations (Lancaster, 2002):

$$E\left(\frac{\partial^2 \mathcal{L}_i}{\partial \tilde{\eta}_i \partial \theta}\right) = 0 \tag{12}$$

This can be interpreted as meaning that the gradient (slope) of the log likelihood with respect to the $\tilde{\eta}_i$ is independent of the gradient of the log likelihood with respect to the common parameter θ.[19]

The purpose of seeking a reparameterization that results in information orthogonality is that it means we can place priors on the parameters and integrate out the fixed-effects. This gives us the marginal posterior for the remaining parameters and we can use Monte Carlo methods to sample values from the marginal posterior to produce estimates and credible intervals for the parameters. The estimates are consistent as $N \to \infty$. The approach relies on the Bayesian idea of integrating out the $\tilde{\eta}_i$ to give us the marginal posterior distribution for the remaining parameters—although, this has many similarities to the frequentist idea of a conditional likelihood (Cox and Reid, 1987; Lancaster, 2000). Note that the orthogonal-reparameterization is not necessary in order to place a prior on and integrate out the incidental parameters but the orthogonalization means that the marginal posterior produced by integrating out the incidental parameters does not depend on the choice of prior that we place on them. In other words: we do not require any information about the individual effects.

If we wish to estimate Equation 3, an orthogonal re-parameterization of the fixed-effects, placement of priors on α, β, σ^2, and the $\tilde{\eta}_i$, and integrating out the reparameterized fixed-effects

[19]More formally, the $\tilde{\eta}_i$ are information orthogonal to θ if the $(\tilde{\eta}_i, \theta)$ components of the expected Fisher information matrix are 0.

results in the following posterior density function (Lancaster, 2002):

$$p(\alpha, \beta, \sigma^2|\text{data}) \propto (\sigma^{-N(T-1)}) \times \exp\left(\frac{N}{T}\sum_{t=1}^{T-1}\frac{T-t}{t}\alpha^t - \frac{1}{2\sigma^2}\sum_{t=1}^{T-1}(y_{i,t} - \alpha y_{i,t-1} - \beta x_{i,t})H(y_{i,t} - \alpha y_{i,t-1} - \beta x_{i,t})\right) \times \pi(\alpha, \beta, \sigma^2) \quad (13)$$

H is defined as an operator that subtracts the mean. $\pi(\alpha,\ \beta,\ \sigma^2)$ represents independent, flat priors on each of the parameters, except α for which we use a uniform $(-1, 1)$ prior. Monte Carlo methods can be used to draw sample values of the parameters from this density distribution and the sampled values can be used to produce point estimates and credible intervals (Appendix D provides details). This approach provides a method of estimating the LRE that differs from other estimators. Instead of dividing the point estimates for β by one minus the point estimate for α, the LRE can be calculated for each pair of sampled β and α values, and the distribution of these LREs used to provide the LRE point estimate and credible interval. Pickup *et al.* (2017) describes the mechanics of using our R package for this estimator.

4.2 The QML approach

Although the MLE of model Equation 3 with a fixed T leads to an incidental parameters problem, one can apply MLE to the first difference version of the model, as used in the GMM estimator Equation 6. Following Hsiao *et al.* (2002), to derive the likelihood for Equation 6, we begin by writing the joint density of $\Delta y_{i,t}$ given $\Delta x_{i,t}$ as:

$$P(\Delta y_{i,t=1:T}|\Delta x_{i,t=1:T}) = P(\Delta y_{i,T}|\Delta y_{i,t=1:T-1}, \Delta x_{i,t=1:T})\cdots P(\Delta y_{i,2}|\Delta y_{i,1}, \Delta x_{i,t=1:T}) \times P(\Delta y_{i,1}|\Delta x_{i,t=1:T})$$

However, the density function is not defined for $\Delta y_{i,1}$ because $y_{i,0}$ is unobserved. Making no assumptions about the distribution of $y_{i,0}$ and treating these as free parameters to be estimated, reintroduces the incidental parameters problem. There are as many of these free parameters as there are cases. To avoid this problem, it is necessary to assume a functional form for the distribution of $\Delta y_{i,1}$. Hsiao *et al.* (2002) propose the following feasible representation:

$$\Delta y_{i,1} = b + \sum_{s=1}^{T}(\Delta x'_{i,s}\pi_s) + \nu_{i,1} \quad (14)$$

where b is a parameter to be estimated and π_s is a $T\times 1$ vector of parameters to be estimated. The derivation of this representation makes assumptions about the covariates (e.g., $x_{i,t}$). It is assumed they are trend—or first difference—stationary, and weakly exogenous. Violations of either assumption will cause biased estimates. Violations of the stationarity assumption can be addressed through variable transformations but endogeneity is an assumption about the DGP. The unmodelled part of $y_{i,t}$ can be a function of current and past values of $x_{i,t}$ but not future values. Hsiao *et al.* (2002) provide the log-likelihood function based on Equations 6 and 14, which can be maximized iteratively (Kripfganz, 2016).[20] Because of the iterative nature of this estimator, convergence on a solution is not guaranteed. This problem does not occur for the

[20]The QML-FE estimation approach is similar to that proposed by Bhargava and Sargan (1983) for a dynamic random-effects model (QML-RE). Unlike Hsiao et al., Bhargava and Sargan do not use a transformed version of

OPM approach, as a closed form expression of the posterior is available and so we do not need to use an iterative procedure to obtain estimates. It is also the case that the dimension of π_s increases with T, resulting in additional parameters to be estimated. There is no equivalent set of parameters for OPM, meaning it may be more efficient than QML-FE as T increases.

4.3 Trade-off between GMM and likelihood-based estimators

GMM is efficient to the extent that valid moment conditions are included and the optimal weighting matrix can be estimated. However, not all moment conditions may be known to the researcher (Kiviet *et al.*, 2017), and estimating the weighting matrix can be computationally difficult. Likelihood-based estimators contain all the information regarding moment conditions in the assumptions made regarding the distribution of the data when constructing the likelihood. Therefore, likelihood-based estimators have an advantage with regards to smaller random errors and smaller confidence intervals/standard errors. A disadvantage of likelihood-based estimators is that if the distributional assumptions are incorrect, the estimator may not be consistent. With GMM estimators, the distributional assumptions are only as strict as the moment conditions included. GMM estimators, as they are applied to panel data, relax assumptions about the distribution by trading off greater random errors, particularly when T and/or N is very small. The value of the trade-off, one way or the other, is unclear as the robustness of the OPM and QML-FE estimators to violations of distributional assumptions is unknown. Their finite sample properties are only partially known and it is unknown how well they perform on tasks typical to political science: estimating the LRE and SRE of covariates; and testing hypotheses regarding these covariates.

5 Estimator performance

5.1 General properties

To evaluate the performance of the TLEs, relative to other estimators, we conduct a series of Monte Carlo experiments. We apply the estimators of interest to data generated by known DGPs. For each DGP we simulate 1000 data sets and evaluate the overall performance of each approach. We begin with the DGP for which GMMs were designed as a solution to the problems encountered when using OLS-FE and GLS-RE:

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \eta_i + \epsilon_{1,i,t} \tag{15}$$

$$x_{i,t} = 0.75\eta_i + \epsilon_{2,i,t}$$

$$\epsilon_{1,i,t} \sim N(0, 1) \quad \epsilon_{2,i,t} \sim N(0, 4) \quad \eta_i \sim U(-\omega, \omega)$$

We use a DGP with a LDV but not a lag of the independent variables because lagged independent variables do not introduce problems for estimation beyond those of contemporaneous independent variables. We also assume the DGP for the independent variable ($x_{i,t}$) contains serial correlation due to the fixed-effects but is not otherwise autoregressive. Appendix E includes simulation results from the inclusion of an autoregressive DGP for $x_{i,t}$. This appendix also tests the consequences of including time-specific effects in the DGP, which we have excluded from Equation 15 for the sake of convenience. Finally, this appendix includes simulations when the true value of α is 0. These variations in the DGP do not change the results for the TLEs reported below.

Equation 3 but when building the likelihood function, it is necessary to have a representation for $y_{i,0}$ and they propose a feasible representation similar to Equation 14. As it is not a TLE, the QML-RE is beyond the scope of this paper.

To explore the efficiency problems of GMM as $T \to 3$ and the potential bias when $\alpha \to 1$, we generate data sets with an N of 1000 and a T of 3, 4 and 5 (Appendix E contains results for $T = 10$), an α of 0.5 and 0.9 and a ω of 1.15 (for a variance of 0.44) and 5 (variance of 8.3). We use $\beta = 0.5$. The $y_{i,t}$ and $x_{i,t}$ are both a function of the individual effects (η_i), with $x_{i,t}$ containing 75 percent of the effect for $y_{i,t}$.

For each data set, we estimate β (i.e., the SRE), α, and the LRE using the following estimators: GMM, QML-FE, and OPM. We also include OLS-FE and GLS-RE to demonstrate the problems that these common estimators encounter when T is very small. We specifically use three GMM estimators. The first is the difference estimator using Equation 6 (Arellano and Bond, 1991) and the second and third are the system estimator discussed earlier (Blundell *et al.*, 2000). These are the estimators outlined and denoted GMM-DIFF and GMM-SYS in Blundell and Bond (1998).[21] For the difference estimator (GMM-Diff), we use all moment conditions available, assuming the exogeneity of $x_{i,t}$. This includes lags 2 through $T-1$ of y_{it} and lags 1 through $T-2$ of x_{it} as instruments in the difference equation. The first system estimator (GMM-Sys1) adds a lag of the first difference of y_{it} in the level equation. The second system estimator (GMM-Sys2) adds to this a lag of the first difference of x_{it} in the level equation. These specifications are common in our literature review. Each GMM estimator uses a two-step procedure which uses parameter estimates from a non-efficient GMM in the first step to estimate the optimal weighting matrix (Arellano and Bond, 1991). This is known to increase efficiency (Roodman, 2009). However, it is also known to produce a downward bias in the estimated standard errors in small samples and so we use the standard error correction proposed by Windmeijer (2005). Given the DGP used for the simulations, these moment conditions are all valid. As a test, we apply a Sargan test to the system estimators. In each set of simulations, we reject the null hypothesis that the instruments (moment conditions) are valid in approximately 5 percent of the cases when using a 0.05 rule. This is the expected false rejection rate, if the instruments are valid. Deleting the cases for when the null hypothesis is rejected makes little difference to the results (Appendix E).

Table 1 presents the results from the Monte Carlo experiment. The mean biases for β and α are both reported. The α coefficient is important because β is only the estimate of the SRE and the full LRE is a function of α: $\beta/(1-\alpha)$. The division by $(1-\alpha)$ means that small errors in α can translate into large errors in the estimated LRE. To demonstrate this, Table 1 also reports the LRE bias. Note that because the estimates of α and β are not independent, the bias in the LRE cannot be determined simply by dividing the β bias by one minus the α bias. Also note that because the LRE is a nonlinear combination of parameters, its distribution is not necessarily symmetric and so we report the median bias in place of mean bias. The root mean squared error for β, α, and the LRE are also reported. The RMSE is an indicator of efficiency and is important because even if an estimator produces the correct answer on average, if it produces a wide range of values, many falling far from the truth, it may be of less value to us than an estimator with a small bias but a narrower range of values. For the LRE, we also report the median deviation from the median estimate (MD). The table reports the biases, RMSEs, and MDs as proportions of the true value. The absolute values are reported in Appendix E. Finally, the proportion of 95 percent confidence intervals that include the true value (the coverage probability statistic) and exclude 0 (the power statistic) are reported. An unbiased estimator that produces accurate standard errors should include the true value within this confidence interval 95 percent of the time and ideally exclude 0, when that is not the true value, 100 percent of the time. These measures reflect the value of the estimator for testing hypotheses regarding the parameter values.

We begin with a summary of the problems encountered by OLS-FE and GLS-RE. The α bias for each is substantial. For OLS-FE, it ranges from 36 to 69 percent of the true value of the

[21] We use the Stata commands xtabond for the difference estimator and xtdpdsys for the system estimators (StataCorp, 2017).

Table 1. Large N properties ($N = 1000$)

T	α	ω	Estimator	α bias	α RMSE	β bias	β RMSE	LRE bias	LRE RMSE	LRE MD	α cov.	β cov.	LRE cov.	α pow.	β pow.	LRE pow.
5	0.9	1.15	OLS-FE	−0.36	0.36	−0.15	0.15	−0.8	0.8	0.0068	0	0	0	1	1	1
5	0.9	1.15	GLS-RE	0.07	0.07	0.022	0.026	1.8	1.8	0.08	0	0.72	0	1	1	1
5	0.9	1.15	GMM-Diff	0.033	0.14	0.015	0.066	0.44	72	0.3	0.95	0.95	0.72	1	1	0.13
5	0.9	1.15	GMM-Sys1	−0.0013	0.03	0.0015	0.022	−0.04	0.34	0.19	0.91	0.96	0.95	1	1	0.99
5	0.9	1.15	GMM-Sys2	−0.0077	0.029	−0.044	0.28	−0.13	0.54	0.28	0.89	0.97	0.96	1	0.94	0.86
5	0.9	1.15	QML	−0.0010	0.028	−0.00058	0.022	−0.017	0.9	0.18	0.95	0.94	0.92	1	1	0.95
5	0.9	1.15	OPM	0.0024	0.03	0.0011	0.022	0.024	0.5	0.19	0.9	0.93	0.89	1	1	1
4	0.9	1.15	OLS-FE	−0.48	0.48	−0.2	0.2	−0.84	0.84	0.0056	0	0	0	1	1	1
4	0.9	1.15	GLS-RE	0.07	0.07	0.022	0.028	1.8	1.8	0.094	0	0.79	0	1	1	1
4	0.9	1.15	GMM-Diff	0.026	0.24	0.012	0.11	0.64	8.6	0.34	0.95	0.95	0.64	0.97	1	0.077
4	0.9	1.15	GMM-Sys1	0.0017	0.042	0.0026	0.028	−0.032	3.4	0.22	0.93	0.96	0.96	1	1	0.84
4	0.9	1.15	GMM-Sys2	−0.0047	0.039	−0.032	0.42	−0.15	3.4	0.36	0.91	0.98	0.96	1	0.72	0.48
4	0.9	1.15	QML	−0.0031	0.044	−0.0013	0.030	0.0033	35	0.25	0.94	0.94	0.91	1	1	0.70
4	0.9	1.15	OPM	0.0029	0.042	0.00062	0.028	0.018	0.86	0.28	0.87	0.92	0.86	1	1	1
3	0.9	1.15	OLS-FE	−0.69	0.69	−0.32	0.32	−0.9	0.9	0.0046	0	0	0	1	1	1
3	0.9	1.15	GLS-RE	0.07	0.07	0.024	0.034	1.7	1.8	0.11	0	0.8	0	1	1	1
3	0.9	1.15	GMM-Diff	0.078	6.1	0.034	2.8	0.86	8.4	0.34	0.96	0.96	0.48	0.58	0.98	0.065
3	0.9	1.15	GMM-Sys1	0.024	0.11	0.01	0.056	0.036	11	0.36	0.99	1	0.93	0.95	1	0.18
3	0.9	1.15	GMM-Sys2	−0.0024	0.073	−0.11	1.3	−0.13	240	0.62	0.98	1	0.99	0.9	0.2	0.062
3	0.9	1.15	QML	−0.0082	0.087	−0.0052	0.05	0.18	26	0.3	0.96	0.95	0.86	1	1	0.29
3	0.9	1.15	OPM	−0.0078	0.06	−0.0032	0.042	−0.0036	0.96	0.38	0.92	0.95	0.92	1	1	1
3	0.5	1.15	OLS-FE	−0.92	0.92	−0.24	0.24	−0.6	0.6	0.015	0	0	0	0.31	1	1
3	0.5	1.15	GLS-RE	0.34	0.34	0.084	0.086	0.66	0.66	0.04	0	0.068	0	1	1	1
3	0.5	1.15	GMM-Diff	−0.011	0.17	−0.0042	0.052	−0.0079	0.27	0.14	0.98	1	0.98	1	1	0.99
3	0.5	1.15	GMM-Sys1	0.0072	0.094	0.002	0.036	0.0093	0.11	0.077	1	1	0.99	1	1	1
3	0.5	1.15	GMM-Sys2	0.0024	0.11	−0.042	1	−0.094	1.1	0.47	0.99	1	1	0.86	0.2	0.19
3	0.5	1.15	QML	−0.0048	0.09	−0.0026	0.038	0.00068	0.13	0.081	0.97	0.95	0.96	1	1	1
3	0.5	1.15	OPM	0.012	0.1	0.0044	0.04	0.01	0.15	0.089	0.9	0.94	0.9	1	1	1
3	0.9	5	GMM-Diff	0.41	12	0.18	5.6	1	3.8	0.092	0.97	0.97	0.27	0.028	0.21	0.12
3	0.9	5	GMM-Sys1	−0.014	0.43	−0.006	0.2	0.054	46	0.98	0.81	1	0.9	0.77	0.86	0.14
3	0.9	5	GMM-Sys2	−0.051	0.14	−0.24	1.1	−1.2	11	1.2	0.84	0.99	0.95	0.88	0.2	0.16
3	0.9	5	QML	−0.0082	0.087	−0.0052	0.05	0.18	26	0.3	0.96	0.95	0.86	1	1	0.29
3	0.9	5	OPM	−0.0077	0.06	−0.0032	0.04	−0.022	0.96	0.38	0.92	0.94	0.92	1	1	1
3	0.5	5	GMM-Diff	−0.18	1.9	−0.046	0.46	0.15	25	0.34	0.98	0.98	0.9	0.3	1	0.16
3	0.5	5	GMM-Sys1	0.026	0.24	0.0064	0.066	−0.0053	0.28	0.19	0.95	1	0.95	0.74	1	0.77
3	0.5	5	GMM-Sys2	−0.024	0.22	−0.74	1.6	−0.79	2.9	0.71	0.97	0.95	0.94	0.71	0.24	0.24
3	0.5	5	QML	−0.0048	0.09	−0.0026	0.038	0.00068	0.13	0.081	0.97	0.95	0.96	1	1	1
3	0.5	5	OPM	0.012	0.1	0.0044	0.04	0.01	0.15	0.089	0.9	0.94	0.9	1	1	1

parameter. For GLS-RE, it can be as small as 7 percent but can also be as large as 19 percent of the true value. For β (the SRE), the biases are smaller but still up to 32 percent for OLS-FE. Further, biases in α translate into large LRE bias tables for both estimators. It gets up to 1.8 times the true value for GLS-RE.

Beyond bias, the coverage statistics for α, β, and the LRE are always below the 0.95 target for both OLS-FE and GLS-RE, often close to or equal to zero. These problems remain whether $\alpha = 0.5$ or 0.9. In contrast, TLEs outperform OLS-FE and GLS-RE in each and every way. It is to these estimators and GMM we now turn.

We start with $\alpha = 0.9$ and $\omega = 1.15$ and note that the RMSEs for β are greatest for GMM-Diff and GMM-Sys2. The RMSEs increase as T decreases from 28 percent of the true value to 130 percent for GMM-Sys2 and from 6.6 to 280 percent for GMM-Diff. This means this estimator regularly produces SRE estimates that are one to three times too large or small. For the TLEs and the GMM-Sys1 estimator, the biases and RMSEs for β are generally always small, although the OPM RMSEs and biases are always equal to or smaller than the GMM and QML-FE estimators. Larger differences between the estimators are apparent in the estimates of α and therefore the LRE.

As T decreases from 5 to 3 ($\alpha = 0.9$; $\omega = 1.15$), the α bias(RMSE) for GMM-Diff increases from 3.3 percent(14 percent) to 7.8 percent(6 times the true value). This is due to the weakness of the moment conditions available to GMM-Diff when α is close to 1 (Stock *et al.*, 2002) and T is small. Because of the weak instrument problem, GMM-Diff periodically produces a *very* poor estimate for α. Because of the division by $(1-\alpha)$ when estimating the LRE, the result is large biases (44 to 86 percent) and extreme RMSEs (many times the true value). Note that unlike the rest of the simulation statistics, the LRE RMSE can be somewhat sensitive to the initial conditions of the simulations. However, it is only sensitive to the extent that when the RMSE is many times the true value, the choice of initial conditions affects how many times greater (but the RMSE is always very large). Under these circumstances it is useful to look at the MD. This is much smaller. However, this does not negate the fact that GMM-Diff periodically produces very poor estimates. This also sometimes results in low power for the LRE (as low as 6.5 percent).

At $T = 5$, the α and the LRE results for the TLEs and GMM-Sys estimators look much better. The biases are small; the RMSEs are many times smaller than for GMM-Diff; the coverage statistics are around 0.95; and the power is about 1.0 (0.86 for GMM-Sys2). As T is reduced to 4 and then 3, GMM-Sys1 and GMM-Sys2 exhibit only moderate RMSEs for α (up to 11 percent) but this translates into large RMSE (many times the true value) and poor power (as low as 6 percent) for the LRE. QML-FE also has a high RMSE and low-power for LRE. Hsiao *et al.* (2002) notes that QML-FE does not estimate the LRE well when α is greater than 0.8 and both T and N are small. It would appear that this is also the case when N is large and T is small. Importantly, OPM shows no such biases, has relatively small RMSEs and full power for all parameters. Perhaps because of the way it estimates the LRE, OPM appears much more robust to the problems GMM and QML-FE estimators encounter when $T < 5$ and $\alpha \to 1$.

When $\alpha = 0.5$, all GMM and TLE estimators perform well even when $T = 3$, confirming the problems are a result of $\alpha \to 1$. However, when ω is increased, the RMSEs for all GMM parameter estimates increase (sometimes significantly) regardless of α, while the TLEs are unaffected. This is consistent with the finding that GMM can encounter weak instrument problems when the between variance is large relative to the within variance.

Previous work in Political Science (e.g. Wawro, 2002, 47) suggest 4–5 waves of data are preferable for GMM. With $T = 5$, and N is large, GMM-Sys1 (and GMM-Sys2 to a lesser extent) resolves many of the problems GMM-Diff exhibits when α is close to 1. GMM-Sys1 is, so far, the optimal GMM estimator, given the DGP. However, all three GMM estimators used valid moment conditions and pass the Sargan test with the expected false rejection rate. Further, even GMM-Sys1 has poor properties for the LRE when $T < 5$ and α is close to 1. Given one can never know the true value of α in the DGP of real-world data, researchers should consider

OPM when they only have 3 or 4 waves of data. This is not an uncommon situation. Of the studies in our literature review 23 percent of them had just three waves. OPM not only performs better under these circumstances, it performs at least as well as other estimators when $T > 4$ and α is not close to 1.

These conclusions are based on a DGP with a large N but many political science examples of small T panel data do not have a large N. Of the studies in our literature review with $T \leq 10$, 38 percent had an $N < 500$ and in almost half of these, N was less than 50.

5.2 Finite sample properties

GMM is known to have poor finite sample properties (Bond and Windmeijer, 2002) and Hsiao *et al.* (2002) find that QML-FE consistently performs better than GMM when both T and N are small. However, they also find that when $\alpha > 0.8$, QML-FE performs poorly when estimating the LRE. QML-FE sometimes does not even converge under such circumstances. The finite sample properties of OPM are entirely unknown. We repeat our simulations with the same DGP as before but with N = 40, 100, 300, and 500; T = 3 and 6; and α = 0.9. We use $\omega = 5$, making between variance 60 percent greater than within variance—not unusual in political science examples. Appendix E contains the results for $\omega = 1.15$. The results are the same but muted. GMM-Sys1 sometimes ran into problems when optimizing the weighting matrix. When this occurred, the results for that particular data set were dropped from the analysis. In principle, QML-FE could run into convergence problems but it did not during our simulations.

With $T = 6$ and $\alpha = 0.9$, GMM-Diff exhibits increasing bias for α and the LRE, as N decreases from 500 to 40. The bias in α increases from 16 to 44 percent. The bias in the LRE increases from 68 to 84 percent. The LRE RMSE increases from 4.4 to 9.2 times the true value. When $T = 3$, the LRE bias is 100 percent and even at $T = 500$, the RMSEs for all parameters (including β) are many times the true values. The improved performance of GMM-Sys1 and GMM-Sys2 relative to GMM-Diff in estimating α is similar to that demonstrated by Blundell and Bond (1998) but GMM-Sys2 exhibits increasing β RMSEs as T decreases—around 24 percent when $T = 6$ and 110 percent when $T = 3$. Also, as demonstrated by Hsiao *et al.* (2002), the QML-FE RMSEs for α and β are often smaller than they are for the GMM estimators but not consistently for α. The OPM RMSEs are consistently the smallest of all estimators. This efficiency of OPM relative to the GMM estimators as $T\&N \rightarrow 0$ is evident but the starkest differences are in the LRE estimates.

Starting with $T = 6$, $N = 500$, the overall best-performing GMM estimator (GMM-Sys1) exhibits a downward LRE bias of 120 percent. As N decreases to 300, then 100 and then 40, the bias increases to 160 percent, then 240 percent. Meanwhile, the RMSEs are many times the true value, and the power drops from 63 to 31 percent. QML-FE exhibits a similar pattern, although with smaller biases. This is consistent with Hsiao *et al.* (2002). QML-FE also has smaller MDs than the GMMs. OPM biases and RMSEs are consistently many times smaller than even QML-FE, and like QML-FE, the MDs are smaller than the GMMs. Even when $N = 40$, the LRE bias is only 11 percent, the RMSE 66 percent, and the MD 34 percent. Noticeably, the power never drops below 100 percent.

When $T = 3$ and N drops to 500, 300, and then 100, the pattern is similar except more extreme. The GMM RMSEs reach 110 times the true value for the LRE, and the power is about 6 percent. For QML-FE, the LRE bias increases from 34 to 72 percent, the LRE RMSE increases from 10 to 12 times the true value, and the LRE power is about 20 percent. For OPM, the LRE bias and RMSE are much smaller. The LRE bias does increase from 9 to 34 percent but even then the RMSE is only 54 percent and the MD 30 percent. Unlike the other estimators, the power for the OPM LRE always remains at 100 percent.

Overall, we see GMM-Diff has poor finite sample properties when $\alpha \rightarrow 1$. GMM-Sys1 (and GMM-Sys2 to a lesser extent) is an improvement but has problems estimating the LRE, and as

$T\&N \to 0$, it exhibits increasing RMSEs for all parameters. (And note, we have used a moderate between- to within-variance ratio. As this gets larger, the GMM RMSEs for all parameters are expected to increase further, while the TLEs are unaffected.) QML-FE is somewhat more efficient than GMM when estimating α and β under these conditions but runs into the same trouble estimating the LRE. OPM appears to improve on both the QML-FE and GMM estimators by being more efficient and much more robust in estimating the LRE when α is close to 1 (Table 2).

5.3 Violations of distributional assumptions

Up to this point, the DGPs we have used meet the distributional assumptions of TLEs. It is not known how well these estimators perform when these distributional assumptions are violated. Next we use DGPs that violate these assumptions in ways that we might expect to encounter in real-world analyses: outliers in the data; skewed variables; bimodal variables; and heteroskedastic errors.

To test the consequences of outliers, we: (1) use the same DGP as in Equation 15 with $\alpha = 0.9$, and $T = 3$ but we use a t-distribution with 10 degrees of freedom to generate the error term for Y_{it}; and (2) we use a t-distribution with 10 degrees of freedom to generate the error term for Y_{it}, while using a t-distribution with 4 degrees of freedom to generate the error term for X_{it}. These distributions have the very fat tails that would result from outliers. To test the consequences of skewed data, we: (1) use a *beta*(2,9) distribution for Y_{it}; (2) use a *beta*(2,9) distribution for Y_{it} and a *beta* (2,5) distribution for X_{it}; (3) use a $\chi^2(2)$ distribution for Y_{it}; and (4) use a $\chi^2(2)$ distribution for Y_{it} and a $\chi^2(2)$ distribution for X_{it}. To test the consequences of a bimodal distribution, we: (1) use a *beta*(0.9,0.9) distribution for Y_{it}; and (2) use a *beta*(0.9,0.9) distribution for Y_{it} and a *beta*(0.5,0.5) distribution for X_{it}. To test the consequences of heteroskedastic errors, we generate $\epsilon_{1,i,t}$ as $e^{(0.2\times x_{i,t})}N(0,\ 1)$.

In each case (Table 3), OPM is largely unbiased and has good power but there are instances (heteroskedastic and χ^2 distributed errors) when the coverage statistics are as low as 65 percent. QML-FE, on the other hand, sometimes exhibits large LRE bias (60 percent when both Y_{it} and X_{it} are χ^2 distributed) and large RMSEs (although MDs are not large). GMM exhibits such biases less often than QML-FE but the LRE RMSEs are just as large. When OPM exhibits poor coverage statistics, GMM does not but the trade-off is that GMM under these same circumstances has poor power, while OPM does not. For the distributional assumption violations we have identified, it would appear that OPM is largely (but not entirely) robust, while QML-FE and GMM exhibit the same problems of estimating the LRE when $T = 3$ and $\alpha \to 1$, as they did without the violations.

6 Empirical applications

In Appendix F, we apply the GMM, QML-FE, and OPM estimators to three empirical examples, demonstrating the consequences of using the different estimators in commonly used data sets and for published research.

7 Concluding thoughts and limitations

We have outlined and tested TLEs for dynamic panel models of very short panel data ($T \leq 10$). As $N \to 0$, QML-FE is somewhat more efficient than GMM but encounters the same issues of bias and large RMSEs for the LRE when α is close to 1, and $T \to 3$. OPM is an improvement on both estimators. It is robust to the problems encountered by GMM and QML-FE when estimating the LRE. It is large N consistent, typically has the smallest RMSEs, best power, best finite sample properties, and is largely robust to violations of distributional assumptions. In all tests conducted in this study, OPM performs as well as or better than the more commonly used GMM estimator.

Table 2. Finite sample properties

T	N	α	Estimator	α bias	α RMSE	β bias	β RMSE	LRE bias	LRE RMSE	LRE MD	α cov.	β cov.	LRE cov.	α pow.	β pow.	LRE pow.
6	500	0.9	GMM-Diff	0.16	0.3	0.08	0.13	0.68	4.4	0.16	0.91	0.91	0.5	0.89	1	0.083
6	500	0.9	GMM-Sys1	−0.063	0.073	−0.017	0.034	−1.2	48	0.6	0.22	0.9	0.69	1	1	0.63
6	500	0.9	GMM-Sys2	−0.066	0.069	−0.14	0.28	−1.8	7.8	0.6	0.096	0.9	0.46	1	0.99	0.83
6	500	0.9	QML-FE	−0.00046	0.031	-6.8×10^{-5}	0.026	0.016	0.62	0.19	0.95	0.95	0.92	1	1	0.92
6	500	0.9	OPM	0.0017	0.031	0.00044	0.026	0.015	0.5	0.2	0.9	0.94	0.91	1	1	1
6	300	0.9	GMM-Diff	0.26	0.39	0.11	0.17	0.76	4.6	0.13	0.9	0.9	0.42	0.76	0.99	0.071
6	300	0.9	GMM-Sys1	−0.069	0.076	−0.019	0.042	−1.6	34	0.82	0.18	0.91	0.74	1	1	0.56
6	300	0.9	GMM-Sys2	−0.07	0.072	−0.11	0.26	−2	7.6	0.66	0.064	0.92	0.44	1	0.98	0.8
6	300	0.9	QML-FE	−0.00056	0.04	−0.0017	0.034	0.02	3	0.22	0.95	0.94	0.91	1	1	0.76
6	300	0.9	OPM	0.0011	0.038	0.0016	0.034	−0.011	0.7	0.24	0.92	0.94	0.91	1	1	1
6	100	0.9	GMM-Diff	0.4	0.54	0.18	0.26	0.82	2.4	0.1	0.84	0.83	0.31	0.52	0.95	0.051
6	100	0.9	GMM-Sys1	−0.074	0.08	−0.013	0.064	−2	32	0.92	0.14	0.94	0.74	1	1	0.51
6	100	0.9	GMM-Sys2	−0.076	0.077	−0.066	0.24	−2.4	76	0.68	0.046	0.94	0.48	1	0.98	0.76
6	100	0.9	QML-FE	−0.0089	0.082	−0.0046	0.06	0.14	34	0.32	0.92	0.95	0.84	1	1	0.4
6	100	0.9	OPM	−0.0044	0.054	−0.0024	0.054	−0.0048	0.84	0.36	0.94	0.95	0.94	1	1	1
6	40	0.9	GMM-Diff	0.44	0.6	0.2	0.28	0.84	9.2	0.096	0.83	0.82	0.3	0.44	0.92	0.059
6	40	0.9	GMM-Sys1	−0.077	0.08	−0.009	0.1	−2.2	17	0.98	0.12	0.93	0.76	1	1	0.53
6	40	0.9	GMM-Sys2	−0.077	0.078	−0.04	0.24	−2.4	56	0.72	0.016	0.93	0.45	1	0.96	0.78
6	40	0.9	QML-FE	−0.021	0.13	−0.013	0.1	0.32	88	0.34	0.86	0.93	0.74	1	1	0.29
6	40	0.9	OPM	−0.022	0.069	−0.0096	0.084	−0.11	0.66	0.34	0.97	0.95	0.97	1	1	1
3	500	0.9	GMM-Diff	1.6	34	0.66	15	1	3.4	0.078	0.98	0.99	0.27	0.015	0.17	0.11
3	500	0.9	GMM-Sys1	−0.013	0.6	−0.006	0.26	−0.092	28	1.3	0.8	1	0.93	0.8	0.86	0.094
3	500	0.9	GMM-Sys2	−0.058	0.16	−0.13	1.2	−1.4	28	1.5	0.79	0.99	0.95	0.88	0.17	0.11
3	500	0.9	QML-FE	−0.013	0.14	−0.0086	0.078	0.34	10	0.34	0.93	0.95	0.78	1	1	0.21
3	500	0.9	OPM	−0.023	0.077	−0.0098	0.054	−0.086	0.76	0.4	0.94	0.94	0.94	1	1	1
3	300	0.9	GMM-Diff	−3.8	110	−1.7	46	1	2.2	0.076	0.98	0.98	0.29	0.012	0.11	0.12
3	300	0.9	GMM-Sys1	−0.041	0.48	−0.016	0.22	−0.3	64	1.4	0.8	1	0.95	0.81	0.88	0.065
3	300	0.9	GMM-Sys2	−0.068	0.14	−0.068	1.1	−1.4	38	1.5	0.8	1	0.96	0.86	0.17	0.1
3	300	0.9	QML-FE	−0.028	0.19	−0.016	0.11	0.44	200	0.46	0.9	0.93	0.73	1	1	0.18
3	300	0.9	OPM	−0.039	0.093	−0.018	0.068	−0.15	0.66	0.38	0.93	0.95	0.93	1	1	1
3	100	0.9	GMM-Diff	0.77	11	0.3	5	1	1.3	0.074	0.99	0.98	0.28	0.007	0.085	0.085
3	100	0.9	GMM-Sys1	−0.037	0.64	−0.0052	0.3	−0.44	110	1.7	0.76	1	0.97	0.83	0.87	0.064
3	100	0.9	GMM-Sys2	−0.07	0.19	−0.026	0.98	−1.5	110	1.5	0.74	1	0.95	0.89	0.16	0.12
3	100	0.9	QML-FE	−0.099	0.33	−0.042	0.19	0.72	12	0.52	0.82	0.89	0.55	1	1	0.21
3	100	0.9	OPM	−0.079	0.13	−0.044	0.12	−0.34	0.54	0.3	0.96	0.93	0.96	1	1	1

Table 3. Violations of distributional assumptions ($T = 3$; $N = 1000$)

DGP	Estimator	α bias	α RMSE	β bias	β RMSE	LRE bias	LRE RMSE	LRE MD	α coverage	β coverage	LRE coverage	α power	β power	LRE power
Outliers														
$Y \sim t(10)$; $X \sim N$	GMM	0.0019	0.1	0.0062	0.052	0.048	110	0.36	0.98	1	0.92	0.97	1	0.19
$Y \sim t(10)$; $X \sim N$	QML-FE	0.0046	0.11	−0.00074	0.06	0.26	16	0.34	0.93	0.94	0.8	1	1	0.27
$Y \sim t(10)$; $X \sim N$	OPM	−0.0033	0.073	−0.0064	0.046	−0.036	1	0.42	0.88	0.92	0.87	1	1	1
$Y \sim t(10)$; $X \sim t(4)$	GMM	0.00044	0.14	0.0092	0.076	0.074	190	0.44	0.97	1	0.89	0.93	0.99	0.15
$Y \sim t(10)$; $X \sim t(4)$	QML-FE	−0.0043	0.16	−0.0019	0.086	0.44	24	0.46	0.87	0.91	0.71	1	1	0.23
$Y \sim t(10)$; $X \sim t(4)$	OPM	−0.0091	0.097	−0.012	0.064	−0.082	1.2	0.48	0.85	0.9	0.85	1	1	1
Skewed distribution (β)														
$Y \sim B(2,\ 9)$; $X \sim N$	GMM	−0.0067	0.069	−0.0026	0.032	−0.052	0.7	0.3	1	1	1	0.91	0.99	0.077
$Y \sim B(2,\ 9)$; $X \sim N$	QML-FE	−0.00034	0.0077	-5.6×10^{-5}	0.005	−0.003	0.074	0.05	0.94	0.93	0.95	1	1	1
$Y \sim B(2,\ 9)$; $X \sim N$	OPM	0	0.0077	0.0001	0.005	0.0034	0.074	0.05	0.95	0.93	0.94	1	1	1
$Y \sim B(2,\ 9)$; $X \sim B(2,\ 5)$	GMM	−0.072	0.94	−0.03	0.44	−0.7	60	1.6	0.61	1	0.87	0.79	0.85	0.22
$Y \sim B(2,\ 9)$; $X \sim B(2,\ 5)$	QML-FE	0.00059	0.14	−0.00074	0.076	0.4	62	0.4	0.9	0.93	0.76	1	1	0.23
$Y \sim B(2,\ 9)$; $X \sim B(2,\ 5)$	OPM	−0.0039	0.084	−0.011	0.056	−0.038	1.1	0.48	0.87	0.92	0.87	1	1	1
Skewed distribution (χ^2)														
$Y \sim \chi^2(2)$; $X \sim N$	GMM	−0.0029	0.11	0.0044	0.056	−0.01	4.2	0.38	0.99	0.99	0.93	0.95	1	0.15
$Y \sim \chi^2(2)$; $X \sim N$	QML-FE	0.0082	0.11	0.0026	0.058	0.28	58	0.34	0.86	0.9	0.75	1	1	0.37
$Y \sim \chi^2(2)$; $X \sim N$	OPM	−0.0044	0.088	−0.007	0.05	−0.038	1.6	0.5	0.74	0.86	0.74	1	1	1
$Y \sim \chi^2(2)$; $X \sim \chi^2(2)$	GMM	0.0019	0.2	0.0096	0.11	0.08	24	0.44	0.96	0.99	0.89	0.89	0.97	0.14
$Y \sim \chi^2(2)$; $X \sim \chi^2(2)$	QML-FE	−0.01	0.2	−0.00072	0.11	0.6	44	0.68	0.75	0.86	0.63	1	1	0.28
$Y \sim \chi^2(2)$; $X \sim \chi^2(2)$	OPM	−0.0081	0.14	−0.024	0.09	−0.088	1.9	0.64	0.65	0.86	0.65	1	1	1
Bimodal distribution														
$Y \sim B(0.9,\ 0.9)$; $X \sim N$	GMM	−0.0069	0.079	−0.0022	0.036	−0.04	1	0.32	1	1	0.99	0.94	0.99	0.11
$Y \sim B(0.9,\ 0.9)$; $X \sim N$	QML-FE	0.00032	0.019	0.0004	0.013	0.005	0.2	0.13	0.96	0.96	0.95	1	1	1
$Y \sim B(0.9,\ 0.9)$; $X \sim N$	OPM	0	0.019	-2.2×10^{-5}	0.013	−0.00096	0.22	0.13	0.96	0.95	0.95	1	1	1
$Y \sim B(0.9,\ 0.9)$; $X \sim B(0.5,\ 0.5)$	GMM	−0.048	0.49	−0.013	0.24	0.096	98	0.92	0.85	1	0.89	0.76	0.84	0.11
$Y \sim B(0.9,\ 0.9)$; $X \sim B(0.5,\ 0.5)$	QML-FE	0.0042	0.16	−0.0018	0.09	0.44	72	0.4	0.9	0.93	0.73	1	1	0.2
$Y \sim B(0.9,\ 0.9)$; $X \sim B(0.5,\ 0.5)$	OPM	−0.0078	0.076	−0.013	0.064	−0.064	0.74	0.42	0.95	0.94	0.95	1	1	1
Heteroskedasticity														
$\epsilon_{i,t} \sim e^{(0.2 \times x_{i,t})} N(0,\ 1)$	GMM	0.0038	0.12	0.0096	0.068	0.078	13	0.38	0.99	0.99	0.91	0.95	1	0.17
$\epsilon_{i,t} \sim e^{(0.2 \times x_{i,t})} N(0,\ 1)$	QML-FE	−0.0068	0.13	−0.0015	0.072	0.32	60	0.4	0.88	0.89	0.75	1	1	0.27
$\epsilon_{i,t} \sim e^{(0.2 \times x_{i,t})} N(0,\ 1)$	OPM	0.0067	0.091	−0.01	0.06	0.054	1.5	0.6	0.77	0.87	0.78	1	1	1

Further, the empirical applications of these estimators in Appendix F demonstrate that the TLEs are different from those of the estimators traditionally used in political science in substantively meaningful ways. That said, it must be kept in mind that we have only been able to explore a finite number of possible data generating processes. Also, while we have attempted to use the most common and best performing (difference and system) GMM estimators, there are other GMM estimators out there that might perform differently. We are not stating that GMM estimators should be abandoned for OPM or QML-FE estimators. We do believe that if a researcher has only three or four waves of data and/or 500 or fewer cases, they should seriously consider the use of OPM as an estimator.

The greatest limitation of the TLEs is the assumption that effects are homogenous across cases (also true of the other estimators considered here). Unfortunately, while basic solutions such as interaction effects can be included, all advanced solutions (to date) require a large T to estimate individual equations for each unit or subsets of units to at least some degree (Pesaran, 2015). Therefore, the estimators we have discussed are appropriate when the researcher is more concerned with the autoregressive nature of the data and individual-effects than with heterogeneous effects. This does not mean they have to believe effects do not vary across individuals. It does mean they must be aware that to the extent that the effects do vary, these estimators will deviate from the true DGP. Of course, this is true of every empirical model and the issue is more about by how much than if it does. Accounting for cross case heterogeneity with a small T is an important avenue for future research, whether the solution is a TLE or some other approach. In the meantime, OPM clearly provides improvements upon existing estimators for small T dynamic models that include fixed-effects.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psrm.2020.30.

Acknowledgments. The authors would like to thank Paul Gustafson, Chris Muris and Vera Tröger for providing feedback on manuscript drafts and the journal editors and reviewers for their constructive comments.

References

Anderson T and Hsiao C (1982) Formulation and estimation of a dynamic models using panel data. *Journal of Econometrics* **18**, 47–82.

Arellano M and Bond S (1991) Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies* **58**, 277–297.

Arellano M, Bond S and Bover O (1995) Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* **68**, 29–51.

Balestra P and Nerlove M (1966) Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas. *Econometrica* **34**, 585–612.

Beck N and Katz J (1996) Nuisance vs. substance: specifying and estimating time-series–cross-section models. *Political Analysis* **6**, 1–36.

Beck N and Katz J (2001) Throwing out the baby with the bath water: a comment on green, yoon and kim. *International Organization* **55**, 487–498.

Beck N and Katz JN (2011) Modeling Dynamics in Time-Series–Cross-Section Political Economy Data. *Annual Review of Political Science* **14**, 331–352.

Bell A and Jones K (2015) Explaining fixed effects: random effects modelling of time-series cross-sectional and panel data. *Political Science Research and Methods* **3**, 113–153.

Bhargava A and Sargan J (1983) Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* **51**, 1635–1659.

Blundell R and Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* **87**, 115–143.

Blundell R, Bond S and Windmeijer F (2000) Estimation in dynamic panel data models: improving on the performance of the standard gmm estimator. In Baltagi BH (ed.), *Nonstationary Panels Cointegrating Panels and Dynamic Panels*. New York, NY: Elsevier Press, pp 53–92.

Bond S and Windmeijer F (2002) Projection estimators for autoregressive panel data models. *Econometrics Journal* **5**, 457–479.

Clark T and Linzer D (2015) Should I Use Fixed or Random Effects?. *Political Science Research and Methods* **3**(2), 399–408.
Cox DR and Reid N (1987) Parameter Orthogonality and Approximate Conditional Inference. *Journal of the Royal Statistical Society. Series B (Methodological)* **49**(1), 1–39.
Doran HE and Schmidt P (2006) GMM Estimators with Improved Finite Sample Properties Using Principal Components of the Weighting Matrix, with an Application to the Dynamic Panel Data Model. *Journal of Econometrics* **133**, 387–409
Finkel S (2008) Linear panel analysis. In Menard S (ed.), *Handbook of Longitudinal Research*. New York, NY: Elsevier Press.
Green DP, Kim SY and Yoon DH (2001) Dirty pool. *International Organization* **55**, 441–468.
Hausman J and Taylor W (1981) Panel data and unobservable individual effects. *Econometrica* **49**, 1377–1398.
Hsiao C (2014) *Analysis of Panel Data*, 3rd Edn. New York, NY: Cambridge University Press.
Hsiao C, Pesaran M and Tahmiscioglu A (2002) Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* **109**, 107–150.
Kiviet J, Pleus M and Poldermans R (2017) Accuracy and efficiency of various gmm inference techniques in dynamic micro panel data models. *Econometrics* **14**, 1–54.
Kripfganz S (2016) xtdpdqml: quasi-maximum likelihood estimation of linear dynamic short-t panel data models. *Stata Journal* **16**, 1013–1038.
Lancaster T (2000) The incidental parameter problem since 1948. *Journal of Econometrics* **95**, 391–413.
Lancaster T (2002) Orthogonal parameters and panel data. *The Review of Economic Studies* **69**, 647–666.
Mundlak Y (1978) On the Pooling of Time Series and Cross Section Data. *Econometrica* **46**(1), 69–85.
Nerlove M (1967) Experimental evidence on the estimation of dynamic economic relations from a time series of cross sections. *Economic Studies Quarterly* **18**, 359–382.
Neyman J and Scott E (1948) Consistent estimation from partially consistent observations. *Econometrica* **16**, 1–32.
Nickell SJ (1981) Biases in dynamic models with fixed effects. *Econometrica* **49**, 1417–1426.
Pesaran MH (2015) *Time Series and Panel Data Econometrics*. Oxford, UK: Oxford University Press.
Pickup M, Gustafson P, Cubranic D and Evans G (2017) Orthopanels: an R package for estimating a dynamic panel model with fixed effects using the orthogonal reparameterization approach. *The R Journal* **9**, 60–76.
Plümper T and Troeger V (2019) Not so harmless after all: the fixed-effects model. *Political Analysis* **27**, 21–45.
Roodman D (2009) A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* **71**, 135–158.
Rosenbaum P (2005) Heterogeneity and Causality. *The American Statistician* **59**(2), 147–152.
Shor B, Bafumi J and Keele L (2007) A Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data. *Political Analysis* **15**, 165–181.
StataCorp (2017) *Stata Longitudinal Data/Panel Data Reference Manual Release 15*. College Station, Tx: Stata Press Publication.
Stock J, Wright J and Yogo M (2002) A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* **20**, 518–529.
Wawro G (2002) Estimating dynamic panel data models in political science. *Political Analysis* **10**, 25–48.
Wilson S and Butler D (2007) A lot more to do: the sensitivity of time-series cross-section analyses to simple alternative specifications. *Political Analysis* **15**, 101–123.
Windmeijer F (2005) The finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of Econometrics* **126**, 25–51.
Wooldridge J (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd Edn. Cambridge, MA: MIT Press.
Wooldridge J (2013) *Introductory Econometrics: a Modern Approach*, 5th Edn. Mason, OH: South-Western.
Zhu L (2012) Panel data analysis in public administration: substantive and statistical considerations. *Journal of Public Administration Research and Theory* **23**, 395–428.

Cite this article: Pickup M, Hopkins V (2020). Transformed-likelihood estimators for dynamic panel models with a very small *T*. *Political Science Research and Methods* 1–20. https://doi.org/10.1017/psrm.2020.30