Asymptotic Properties of Empirical Likelihood Estimator in Dynamic Panel Data Models

Günce Eryürük

North Carolina State University[1]

February 1, 2009

[1]Department of Economics, Box 8110, North Carolina State University, Raleigh, NC 27695, USA. Email: geryuru@ncsu.edu

Abstract

Empirical likelihood (EL) is shown to have various advantages over other methods like generalized method of moments (GMM). These advantages are especially important under the many and weak moments settings, which are often used in econometrics. Little is known as yet on the relative merits of these estimators in panel data models, which offer great flexibility to empirical researchers. In this article, we try to fill that gap by establishing the asymptotic properties of the EL estimator for a dynamic panel model with individual effects when both N and T tend to infinity. We give the conditions under which this estimator is consistent and asymptotically normal.

JEL Classification: C13, C23.

Keywords: Dynamic panel data model, empirical likelihood, fixed effects, limited information maximum likelihood, generalized method of moments.

1 Introduction

Dynamic panel data models offer great flexibility to empirical researchers. Many economic phenomena are dynamic in nature. Examples include household consumption, firms' factor demands, and countries' economic growth. These models, while handling this issue, allow researches to control for unobserved heterogeneity in adjustment dynamics between different individual units such as households, firms, or countries. By this way, these models provide an improved insight into those phenomena.

When dynamic models are estimated using panel data, the "usual" least squares methods lead to inconsistent estimates for the parameters of the models when the time dimension, T, is short regardless of the cross sectional dimension, N. This inconsistency stems from the fact that the disturbance terms are correlated with the lagged endogenous variable [1]. Moreover, under large N, fixed T asymptotics it is well known from Nickell (1981) that the standard maximum likelihood estimator suffers from an incidental parameter problem leading to inconsistency. In order to avoid this problem, the literature has focused on generalized method of moments (GMM) estimator applied to first differences. Examples include Anderson & Hsiao (1982), Holtz-Eakin, Newey, & Rosen (1988), and Arellano & Bond (1991).

However, the conclusions of many Monte Carlo studies on the dynamic panel data models show that for these estimators that use the full set of moments available for errors in first differences can be severely biased, specifically when the instruments are weak[2] and the number of moments is large relative to the cross section sample size (Arellano & Bond (1991), Kiviet (1995), Zilliak (1997), Blundell & Bond (1998), and Alonso-Borrego & Arellano (1999)).

[1] The nature and the magnitude of this inconsistency are well defined by Nickell (1981) and Sevestre and Trognon (1985) for fixed and random effects specifications, respectively

[2] This occurs when the series are highly autoregressive, i.e. the autoregressive parameter being close to one, and the relative variance of fixed effects to the variance of idiosyncratic shocks is large.

Many empirical applications show that this low precision of GMM is also evident in models other than the dynamic panel data models. To improve the small sample properties of GMM estimators, a number of alternative estimators have been suggested. Among which include EL, continuous updating (CU), and exponential tilting estimators. Newey & Smith (2004) show that these estimators are members of a class of generalized empirical likelihood (GEL) estimators. They use this structure to compare their higher order asymptotic properties with that of GMM. They find that EL has two theoretical advantages. First, its asymptotic bias does not grow with the number of moment restrictions, while the bias of GMM often does. This, as a result, suggests that in estimation of models with many moment conditions, which are often used in empirical studies, the bias of EL will be less than the bias of GMM and hence EL can be an important alternative to GMM in such applications. Furthermore, they show that under a symmetry condition, which may be satisfied in some instrumental variable settings, all the GEL estimators inherit the small bias property of EL. The second theoretical advantage of EL is that after it is bias corrected, using probabilities obtained from EL, it is higher order efficient relative to other bias corrected estimators.

The purpose of this article is to provide further insight into the asymptotic properties of EL estimator in dynamic panel data framework by allowing both N and T tend to infinity and study its behaviour for alternative relative rates of increase for N and T. This asymptotics is motivated by the increased availability of panel data sets covering different individuals, regions, and countries over a relatively long time period. Among the important examples of these data sets, the PSID household panel in the US, Penn World table and the balance sheet-based company panels that are available in many countries can be considered. For panels in which T is not negligible relative to N, the analysis of the asymptotic behaviour of the estimators as both T and N tend to infinity may provide better approximations to the finite sample behaviour of the estimators and hence may be useful in assessing alternative methods. Previously, this type of asymptotics is used by Alvarez

& Arellano (2003). They derive the asymptotic properties of within groups (WG), GMM, and Limited Information Maximum Likelihood (LIML) estimators.

It is also the case that as T grows, for each period, the number of available lags which can be used as instruments for the equations in the first differences grow at the rate of $T(T-1)/2$. This corresponds to what is known in the literature as "many moment conditions" situation. It is well known, in the linear instrumental variable regression models that using many moments causes the usual Gaussian asymptotic approximation to be poor. In more general contexts, among others, this problem is addressed by Han & Phillips (2006) and Newey & Windmeijer (2007). They point out that the two-step GMM estimator can be very biased. On the other hand, GEL estimator has smaller bias but the usual standard errors are found to be too small. In their study, Newey & Windmeijer (2007), consider alternative asymptotics that addressed this problem, i.e. asymptotic properties of GEL and GMM under "any weak moment conditions". They find that the two-step GMM estimator is asymptotically biased under this scheme, whereas GEL estimator is not. In addition, they find that, under the alternative asymptotics, GEL has a Gaussian limit distribution with asymptotic variance larger that the usual one and for this estimator they propose an appropriate variance estimator that is consistent under standard and alternative asymptotics.

It is natural to expect to find similar advantages of GEL estimators in dynamic panel data models under the many moment setting, as well. Unfortunately, little is known on the relative merits of these estimators in dynamic panel data models under the double asymptotics, i.e. asymptotics taken as both T and N going to infinity. In a Monte Carlo study, Oĝuzoĝlu (2006) compares performance of a number of estimators which includes GMM, EL, transformed ML, minimum distance and bias corrected LSDV estimators in an autoregressive panel model for various parameter combinations. The results show that the bias of all estimators considered tends to increase as the autoregressive parameter get larger. The increase in bias is the highest for LSDV, whereas, EL is the least sensitive to changes in this parameter. Moreover, the bias of GMM does not decrease much as T increases. When the overall performances are concerned, i.e. comparison based on biases, standard deviations, and root mean square errors, EL becomes the best performer.

Gonzalez (2007), in the same framework, considers the finite-sample size properties of the overidentification tests for a hybrid of EL and bootstrap estimators. Previously, a similar study is done by Brown & Newey (2001) and Bowsher (2002) for the GMM estimator. Gonzalez (2007) investigates whether the limitations encountered within GMM estimation are extended to EL-bootstrap estimator. Her results show that EL-bootstrap may be a good alternative to GMM estimator within this setting. She also applies this estimator using the cash-flow series data for 174 US firms.

Although a few studies have been done considering the finite sample performance of the EL estimator in dynamic panel data models, none of them, to our knowledge, analyze its asymptotic performance explicitly under the aforementioned settings. This article tries to fill this gap. Specifically, we establish the asymptotic properties of the EL estimator for a first-order autoregressive model with individual effects when both N and T tend to infinity. We show that this estimator is consistent and asymptotically normal. We also compare the asymptotic properties of EL estimator with those of the GMM and LIML estimators, which are popular in empirical research.

The paper is organized as follows. Section 2 presents the model and the estimators. In section 3 we establish the asymptotic properties of the EL estimator. For comparison purposes we give those for the LIML and GMM estimators, as well. A comparison of these estimators in finite samples using Monte Carlo simulations is given in section 4. Section 5 concludes and states plans for future work. The Appendix gives the proofs.

2 The Model and The Estimators

2.1 The Model

We consider a first order univariate autoregressive panel data model given by

$$(1) \qquad y_{it} = \alpha_0 y_{i,t-1} + \eta_i + v_{it}, \quad \text{for} \quad t = 1, \dots, T; \;\; i = 1, \dots, N$$

where y_{it} is the observable variable whose dynamics are of interest; for example, local government expenditure variable, $|\alpha_0| < 1$, η_i is the fixed effect representing the unobserved heterogeneity among individuals, and v_{it} is the idiosyncratic variable with zero mean and variance σ^2 given $\eta_i, y_{i0}, \dots, y_{i,t-1}$ and has no autocorrelation. We assume that y_{i0} is observed. Define $x_{it} \equiv y_{i,t-1}$.

The parameter of interest is α_0. Our goal is to analyze the asymptotic properties of EL estimator of this parameter. For comparison purposes we are going to consider that of GMM and LIML estimators. Next we shall define these estimators.

2.2 The Estimators

The GMM Estimator. The GMM estimator considered here is a version developed by Arellano & Bover (1995), which simplifies characterization of the "weight matrix" in GMM estimation. Arellano & Bover (1995) eliminate the fixed effect η_i in (1) by applying Helmert's transformation. For example, the t-th element of transformed v_{it} can be written as:

$$v_{it}^* = c_t[v_{it} - \frac{1}{T-t}(v_{i,t+1} + \dots + v_{iT})] \qquad t = 1, \dots, T-1$$

where $c_t^2 = (T-1)/(T-t+1)$. That is, to each of the first $(T-1)$ observations the mean of the remaining future observations available in the sample is subtracted. The weighting c_t is introduced to equalize the variances. This transformation can be applied by using the forward orthogonal deviations operator, A, where

$$A = diag[\frac{T-1}{T}, \ldots, \frac{1}{2}]^{1/2} \begin{bmatrix} 1 & -\frac{1}{T-1} & -\frac{1}{T-1} & \ldots & -\frac{1}{T-1} & -\frac{1}{T-1} & -\frac{1}{T-1} \\ 0 & 1 & -\frac{1}{T-2} & \ldots & -\frac{1}{T-2} & -\frac{1}{T-2} & -\frac{1}{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \ldots & 0 & 1 & -1 \end{bmatrix}.$$

Equation (1) with the variables stacked over t can be written as

$$y_i = \alpha_0 x_i + \eta_i \imath_T + v_i$$

where $y_i = [y_{i1}, \ldots, y_{iT}]'$, $x_i = [x_{i1}, \ldots, x_{iT}]'$, $v_i = [v_{i1}, \ldots, v_{iT}]'$, and $\imath$ dimension T vector of ones. Operating A on this equation produces the transformed model:

(2) $$y_i^* = \alpha_0 x_i^* + v_i^*$$

where $y_i^* = Ay_i$, $x_i^* = Ax_i$, $v_i^* = Av_i$. Note that the fixed effect are eliminated because $A\imath = 0$. Also, $A'A = Q_T \equiv I_T - \imath_T \imath_T'/T$ (Q_T is known as WG operator) and $AA' = I_{T-1}$. Thus, if $Var(v_i) = \sigma^2 I_T$, the vector of errors in orthogonal deviations also has $Var(v_i^*) = \sigma^2 I_{T-1}$.

Let $z_{it} = [x_{i1}, \ldots, x_{it}]'$. The model (2) and the stated conditions imply the following moment conditions

(3) $$E[z_{it} v_{it}^*] = 0 \qquad t = 1, \ldots, T-1.$$

There are $m \equiv T(T-1)/2$ orthogonality conditions. These moment conditions can be written, more compactly, as

$$E[Z_i' v_i^*] = 0,$$

where

$$\underset{(T-1)\times q}{Z_i} = \begin{bmatrix} \underset{1\times 1}{z_{i1}}' & 0 & \dots & 0 \\ 0 & \underset{1\times 2}{z_{i2}}' & 0 \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \underset{1\times (T-2)}{z_{iT-1}}' \end{bmatrix} = \begin{bmatrix} y_{i0} & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & y_{i0} & y_{i1} & & 0 & & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & 0 & \dots & y_{i0} & \dots & y_{i(T-2)} \end{bmatrix}.$$

The constant variance of v_{it} given $\eta_i, y_{i0}, \dots, y_{i,t-1}$ implies that

$$E(Z_i' v_i^* v_i^{*\prime} Z_i) = \sigma^2 E(Z_i' Z_i). \tag{4}$$

Therefore, letting $x^* = (x_1^{*\prime}, \dots x_N^{*\prime})'$ and $y^* = (y_1^{*\prime}, \dots y_N^{*\prime})'$, an asymptotically efficient GMM estimator of α_0 based on the moment conditions in (3) is given by

$$\widehat{\alpha}_{GMM} = \frac{x^{*\prime} Z (Z'Z)^{-1} Z y^*}{x^{*\prime} Z (Z'Z)^{-1} Z x^*}$$

where $Z = (Z_1', \dots, Z_N')'$.

The LIML Estimator. A non-robust analog of the LIML estimator of the simultaneous equations literature solves the following problem:

$$\widehat{\alpha}_{LIML} = \arg\min_{\alpha} \frac{(y^* - \alpha x^*)' Z (Z'Z)^{-1} Z' (y^* - \alpha x^*)}{(y^* - \alpha x^*)'(y^* - \alpha x^*)}.$$

The robust LIML analog, or "continuously updated" GMM estimator in the terminology of Hansen, Heaton & Yaron (1996), can be written as

$$\widehat{\alpha}_{CU} = \arg\min_{\alpha} (y^* - \alpha x^*)' Z \left(\sum_{i=1}^{N} Z_i' (y_i^* - \alpha x_i^*)(y_i^* - \alpha x_i^*)' Z_i \right)^{-1} Z' (y^* - \alpha x^*).$$

In the non-robust version, instead of keeping σ^2 fixed in the weighting matrix of the GMM criterion, it is continuously updated by making it a function of the arguments in the estimating criterion.

The EL Estimator. Empirical Likelihood estimation (Qin & Lawless (1994) and Imbens (1997)) is a one-step method that achieves the same first-order asymptotic efficiency as robust GMM.

The empirical likelihood estimator maximizes a multinomial pseudo likelihood (or empirical likelihood) function subject to the orthogonality conditions. Letting p_i be the probability of observation i, the multinomial log likelihood of the data is given by the empirical likelihood estimator:

$$L = \sum_{i=1}^{N} \ln p_i.$$

The EL estimator maximizes this function subject to the restrictions

$$p_i \geq 0, \sum_{i=1}^{N} p_i = 1 \text{ and } \sum_{i=1}^{N} p_i Z_i'(y_i^* - \alpha x_i^*) = 0.$$

The Lagrangian is given by

$$\mathcal{L} = \sum_{i=1}^{N} p_i + \phi(1 - \sum_{i=1}^{N} p_i) - N\lambda' \sum_{i=1}^{N} p_i Z_i'(y_i^* - \alpha x_i^*),$$

where λ and ϕ are Lagrange multipliers. Taking the derivative of $\mathcal{L}$ with respect to p_i we obtain the following first-order conditions

$$\frac{1}{p_i} - \phi - N\lambda' Z_i'(y_i^* - \alpha x_i^*) = 0.$$

Multiplying by p_i and adding equations we get $\phi = N$. Hence,

$$p_i = \frac{1}{N}\left(\frac{1}{1 + \lambda' Z_i'(y_i^* - \alpha x_i^*)}\right).$$

The multipliers of the moment conditions can be determined as implicit functions $\lambda(\alpha)$ solving (for a given value of α):

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{1 + \lambda' Z_i'(y_i^* - \alpha x_i^*)}\right) Z_i'(y_i^* - \alpha x_i^*) = 0$$

such that $1 + \lambda' Z_i'(y_i^* - \alpha x_i^*) \geq 1/N$.

The concentrated likelihood function for α:

$$\mathcal{L}_c(\alpha) = \prod_{i=1}^{N} \frac{1}{N}\left(\frac{1}{1 + \lambda' Z_i'(y_i^* - \alpha x_i^*)}\right).$$

Therefore, the EL estimator is given by

$$\widehat{\alpha}_{EL} = \arg\min_{\alpha} \sum_{i=1}^{N} \ln[1 + \lambda(\alpha)' Z_i'(y_i^* - \alpha x_i^*)].$$

A computationally useful alternative expression for α_{EL} is

$$\widehat{\alpha}_{EL} = \arg\min_{\alpha} \widehat{Q}(\alpha), \text{ where } \widehat{Q}(\alpha) = \max_{\lambda} \frac{1}{NT} \sum_{i=1}^{N} \ln[1 + \lambda' Z_i'(y_i^* - \alpha x_i^*)].$$

3 The Asymptotic Properties Of The Estimators

In this section we derive the asymptotic properties of the previous estimators when both N and T tend to infinity. Following Alvarez & Arellano (2003), we make the following assumptions:

Assumption 1. $\{v_{it}\}$ $(t = 1, \ldots T; i = 1, \ldots N)$ *are i.i.d across time and individuals and independent of* η_i *and* y_{i0} *with* $E[v_{it}] = 0$, $Var[v_{it}] = \sigma^2$, *and finite moments up to fourth order.*

Assumption 2. *The initial observations satisfy*

$$y_{i0} = \frac{\eta_i}{1 - \alpha_0} + \omega_{i0}$$

where ω_{i0} *is independent of* η_i *and i.i.d. with the steady state distribution of the homogenous process, so that* $\omega_{i0} = \sum_{j=0}^{\infty} \alpha_0^j v_{i,(-j)}$.

Assumption 3. η_i *are i.i.d. across individuals with* $E[\eta_i] = 0$, $Var[\eta_i] = \sigma_\eta^2$ *and finite fourth order moment.*

Note that under these assumptions, the moment conditions given in (3) do not represent all the available moment conditions available. Ahn & Schmidt (1995) present additional moment conditions and argue that they are important in improving the GMM estimation in highly persistent samples. However, we focus only on the moment conditions in (3) as they remain valid under much weaker assumptions.

3.1 The GMM and the LIML Estimators

Alvarez & Arellano (2003) show that under the stated assumptions as both N and T tend to infinity, provided $(\log T)/N \to 0$, $\widehat{\alpha}_{GMM}$ is consistent for α_0:

$$\widehat{\alpha}_{GMM} \xrightarrow{p} \alpha_0$$

Moreover, provided $T/N \to c$, $0 \leq c \leq \infty$

$$\sqrt{NT}\left[\widehat{\alpha}_{GMM} - \left(\alpha_0 - \frac{1}{N}(1+\alpha_0)\right)\right] \xrightarrow{d} \mathcal{N}\left(0, 1-\alpha_0^2\right).$$

Although, number of moment conditions, m tend to infinity (at the rate T^2), their result show that $\widehat{\alpha}_{GMM}$ remains consistent. However, in the structural equation setting, when both the number of instruments and the sample size tend to infinity, while their ratio tends to a positive constant, the two-stage least squares estimator is shown to be inconsistent (Kunitomo (1980), Morimune (1983), and Bekker (1994)). The intuition for this consistency of $\widehat{\alpha}_{GMM}$ is defined by Alvarez & Arellano (2003) as in structural equation setting too many instruments produces over fitting and undesirable closeness to the OLS coefficients. Here a large number of instruments is associated with larger values of T and in such a case closeness to OLS, which is the WG estimator, becomes increasingly desirable because "endogeneity bias" tends to zero as $T \to \infty$.

For, LIML estimator, they show that, under the stated assumptions, as both N and T tend to infinity, provided $T/N \to c$, $0 \leq c \leq 2$, $\widehat{\alpha}_{LIML}$ is consistent for α_0:

$$\widehat{\alpha}_{LIML} \xrightarrow{p} \alpha_0$$

Moreover,

$$\sqrt{NT}\left[\widehat{\alpha}_{LIML} - \left(\alpha_0 - \frac{1}{2N-T}(1+\alpha_0)\right)\right] \xrightarrow{d} \mathcal{N}\left(0, 1-\alpha_0^2\right).$$

Note that GMM and LIML estimators are both asymptotically normal with the same asymptotic variance, however, unless $T/N \to 0$, they exhibit a bias term in their asymptotic distributions differing in its order of magnitude: $((1+\alpha)/N$ for GMM and $((1+\alpha)/(2N-T)$ for LIML. Provided $T < N$, the LIML has a smaller asymptotic bias.

3.2 The EL Estimator

For consistency and asymptotic normality of EL estimator some additional assumptions are needed. Let $\lambda_{min}(S)$ and $\lambda_{max}(S)$ denote the smallest and the largest eigenvalues of a symmetric matrix S, respectively.

Assumption 4. *(i) There is $C > 0$ such that $1/C \leq \lambda_{min}\big(E[Z_i'(y_i^* - \alpha x_i^*)(y_i^* - \alpha x_i^*)'Z_i]\big), \lambda_{max}\left(E[Z_i'(y_i^* - \alpha x_i^*)(y_i^* - \alpha x_i^*)'Z_i]\right) \leq C,$ and $\lambda_{max}\left(E[Z_i'x_i^*x_i^{*\prime}Z_i]\right) \leq C$; (ii) $\sup_{\alpha}\|\frac{1}{N}\sum_{i=1}^{N} Z_i'(y_i^* - \alpha x_i^*)(y_i^* - \alpha x_i^*)'Z_i - E[Z_i'(y_i^* - \alpha x_i^*)(y_i^* - \alpha x_i^*)'Z_i]\| \xrightarrow{p} 0.$*

Assumption 5. *For a constant C and $\gamma > 2$, $E|v_{it}|^{2\gamma} < C$ and $E|\eta_i|^{2\gamma} < C$.*

Assumption 4 puts restrictions on the rate at which T and hence the number of moment conditions, m, can grow relative to N. Assumption 5 puts further moment restriction on the error and the fixed effect terms. Note that Assumption 5 along with Assumption 2 imply that $E|y_{i0}|^{2\gamma} < C$.

The following is a consistency result for the EL estimator.

Theorem 1. *Let Assumptions 1-5 hold. Then as both N and T tend to infinity, provided $N^{1/\gamma}T^{3-2/\gamma}\sqrt{T^2/N} \to 0$, $\gamma > 2$, $\widehat{\alpha}_{EL}$ is consistent for α_0:*

$$\widehat{\alpha}_{EL} \xrightarrow{p} \alpha_0.$$

For the consistency of the EL estimator a further restriction is need on the relative rates at which T and N can grow. This is also the case in Newey & Windmeijer (2007) for a general cross sectional model.

Next we give the asymptotic normality result. This result is parallel to Theorem 3 of Newey & Windmeijer (2007).

Theorem 2. *Let Assumptions 1-5 hold. Then as both N and T tend to infinity, provided $N^{1/\gamma}T^{3-2/\gamma}\sqrt{T^2/N} \to 0$, $\gamma > 2$, and $T^{11}/N \to 0$,*

$$\sqrt{NT}(\widehat{\alpha}_{EL} - \alpha_0) \xrightarrow{d} N(0, 1 - \alpha_0^2).$$

Note that Newey & Windmeijer (2007) give an asymptotic variance of GEL estimator as a summation of two terms. The first term corresponds to the conventional asymptotic variance term of GMM. The additional term can be considered as a "higher order" variance term in asymptotic theory with fixed number of moment conditions. They note that this term can be important even when the sample size is large under certain conditions, which includes weak moments. Under the restrictions on the relative rates on N and T given by Theorem 2, the additional terms, in our case, tend to zero[3]. Hence, the gradient of the EL objective function at the true parameter, α_0, scaled by $\sqrt{NT}$ takes the following form:

$$\sqrt{NT}\frac{\partial\widehat{Q}(\alpha_0)}{\partial\alpha} = \sqrt{NT}\left(\frac{1}{NT}\sum_{i=1}^{N} x_i^{*\prime} Z_i\right)\left(\frac{1}{NT}\sum_{i=1}^{N} Z_i' v_i^* v_i^{*\prime} Z_i\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N} Z_i' v_i^*\right) + o_p(1).$$

The asymptotic variance of $\sqrt{NT}\frac{\partial\widehat{Q}(\alpha_0)}{\partial\alpha}$ converges to[4]

$$\frac{1}{\sigma^2 T} E\Big(x_i^{*\prime} Z_i\Big) E\Big(Z_i' x_i^* x_i^{*\prime} Z_i\Big)^{-1} E\Big(Z_i' x_i^{*\prime}\Big) \to \frac{1}{\sigma^2 T} E\Big(x_i^{*\prime} x_i^*\Big).$$

From Alvarez & Arellano (2003) equation $A8$, we have

$$\frac{1}{\sigma^2 T} E\Big(x_i^{*\prime} x_i^*\Big) = \frac{1}{1-\alpha_0^2} - \frac{1}{T}\frac{\sigma^2}{1-\alpha_0^2}\left(\frac{(1+\alpha_0)}{(1-\alpha_0)} - \frac{1}{T}\frac{2\alpha_0(1-\alpha_0^T)}{(1-\alpha_0)^2}\right).$$

We have independence across i, so by the central limit theorem

$$\sqrt{NT}\frac{\partial\widehat{Q}(\alpha_0)}{\partial\alpha} \xrightarrow{d} N\left(0, \frac{1}{1-\alpha_0^2}\right).$$

Then, as shown in Appendix, theorem 2 follows from theorem 3 of Newey & Windmeijer (2007).

Theorem 2 requires that for asymptotic normality T has to grow much slower than N does. More specifically, it is required that $\lim(T/N) \to c = 0$. This condition is much more strict than $0 \le c \le 2$ requirement which is needed for asymptotic normality of LIML. Moreover, for LIML, when $c = 0$, the asymptotic bias disappears and the two estimators become asymptotically equal.

[3]The proof is available upon request from the author.

[4]For the second convergence see page 13 of Newey & Windmeijer (2007).

In the literature, performances of these estimators are compared through Monte Carlo studies. Although, to our knowledge none of these studies have compared all these three estimators in the same model under the same specifications, they have been compared pairwise on not so different settings. Newey & Windmeijer (2007) consider GMM and GEL estimators in a panel data model with heterogenous idiosyncratic errors and predetermined variables which are introduced via inclusion of a lagged independent variable as an explanatory variable. They compare their interquartile range and median bias. They show that when the instruments are strong[5] biases are negligible for these estimators with comparable interquartile ranges. When the instruments are weaker[6], the GMM estimators are downward biased, whereas the CU estimator is median unbiased but exhibits a larger interquartile range that the GMM estimators. Moreover, when they included lags of dependent variable as sequential instruments additional to the sequential lags of the independent variable the GMM estimators become more downward biased whereas the CU estimator is still median unbiased with an interquartile range decreased by 15%.

Oĝuzoĝlu (2006) compares the performances of EL and GMM estimators, among the others, based on their bias, standard deviation and root mean square, in a model identical to the one we consider here. He shows that in the case of strong instruments the EL estimator performs better in general. In the case of weak instruments, EL estimator, again, exhibits the best performance except when T is small. In this case, its performance suffers mostly from its high variance.

Although, for EL estimator, the condition on the rate at which T can grow is much more strict than that of LIML and GMM for asymptotic normality, Monte Carlo studies suggest that EL usually outperforms LIML or GMM in various settings.

[5]The case in which the coefficient of the lagged independent variable is equal to 0.40.

[6]the case in which the coefficient of the lagged independent variable is equal to 0.85.

4 Monte Carlo Study

In this section we report some Monte Carlo simulations of EL, GMM, and LIML estimators for various combinations of values of T and N. The purpose of these experiments is to compare the biases of these estimators for different values of T and N. For all cases we conducted 1000 replications from the model specified in section 2 and 3 under normality, i.e. each sample consists of N independent observation of $(y_{i0}, y_{i1}, \ldots, y_{iT})$ generated from the process $y_{i0} = (1-\alpha)^{-1}\eta_i + (1-\alpha^2)^{-1/2}v_{i0}$, $y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}$ for $t = 1, \ldots, T$ with $v_i = (v_{i0}, v_{i1}, \ldots, v_{iT})' \sim N(0, I)$ and $\eta_i \sim N(0, \sigma_\eta^2)$ independent of v_i. For all cases σ^2 and σ_η^2 are 1 and 0, respectively. The design of this Monte Carlo experiment follows Alvarez & Arellano (2003) closely.

In Table 1 we report median, interquartile range (iqr), and median absolute error (mae) of the EL, GMM, and LIML estimators for $\alpha = 0.2$, 0.5 and 0.8 and for $N = 100$ with

Table 1: Medians, Interquartile Ranges, and Median Absolute Errors of the Estimators ($N = 100$)

	$\alpha_0 = 0.2$			$\alpha_0 = 0.5$			$\alpha_0 = 0.8$		
	EL	GMM	LIML	EL	GMM	LIML	EL	GMM	LIML
					$T^0 = 10$				
median	0.1949	0.1830	0.1946	0.4908	0.4750	0.4912	0.7955	0.7801	0.8002
iqr	0.0589	0.0577	0.0595	0.0623	0.0575	0.0574	0.0571	0.0533	0.0566
mae	0.0294	0.0292	0.0298	0.0313	0.0285	0.0288	0.0279	0.0258	0.0281
					$T^0 = 25$				
median	0.1870	0.1825	0.1914	0.4841	0.4789	0.4902	0.7824	0.7756	0.7907
iqr	0.0334	0.0306	0.0309	0.0317	0.0295	0.0303	0.0272	0.0241	0.0250
mae	0.0166	0.0153	0.0155	0.0158	0.0149	0.0152	0.0136	0.0120	0.0126
					$T^0 = 50$				
median	0.1852	0.1824	0.1890	0.4830	0.4797	0.4881	0.7804	0.7771	0.7873
iqr	0.0197	0.0197	0.0208	0.0205	0.0193	0.0203	0.0147	0.0135	0.0149
mae	0.0097	0.0098	0.0104	0.0102	0.0098	0.0100	0.0074	0.0068	0.0075

Table 2: Medians, Interquartile Ranges, and Median Absolute Errors of the Estimators ($N = 50$)

	$\alpha_0 = 0.2$			$\alpha_0 = 0.5$			$\alpha_0 = 0.8$		
	EL	GMM	LIML	EL	GMM	LIML	EL	GMM	LIML
					$T^0 = 10$				
median	0.1834	0.1625	0.1834	0.4790	0.4549	0.4843	0.7823	0.7510	0.7894
iqr	0.0837	0.0798	0.0858	0.0945	0.0817	0.0899	0.0855	0.0742	0.0793
mae	0.0418	0.0400	0.0429	0.0470	0.0413	0.0446	0.0423	0.0376	0.0399
					$T^0 = 25$				
median	0.1766	0.1666	0.1814	0.4714	0.4587	0.4778	0.7702	0.7578	0.7824
iqr	0.0492	0.0397	0.0428	0.0487	0.0401	0.0430	0.0390	0.0341	0.0370
mae	0.0245	0.0199	0.0215	0.0246	0.0201	0.0217	0.0195	0.0170	0.0186
					$T^0 = 50$				
Median	0.1689	0.1663	0.1690	0.4643	0.4603	0.4634	0.7658	0.7613	0.7667
iqr	0.0327	0.0276	0.0287	0.0319	0.0263	0.0284	0.0239	0.0194	0.0242
mae	0.0164	0.0138	0.0143	0.0160	0.0131	0.0142	0.0120	0.0097	0.0120

$T^0 = 10$, 25, and 50, where $T^0 = T + 1$ (the actual number of time series observations in the data). Table 2 reports the similar results for $N = 50$.

Table 1 and 2 reveals that in all cases the median bias of GMM estimator is always larger than the median biases of the EL and LIML estimators. The EL and LIML biases are both very small, however, the ranking between the two is not obvious. When T is small relative to N and α_0 is small EL bias is smaller than the LIML bias. The difference between them gets smaller for a square panel with $T^0 = 50$ and $N = 50$.

When it comes to dispersion, GMM always has a smaller interquartile range than the other two estimators. Again, the ranking between EL and LIML is not clear. When $T^0 = 25$ LIML interquartile range is always smaller, however, for the other cases there is not an obvious order.

Finally, for median absolute errors, Table 1 and 2 show that except one case (the case when $T^0 = 50$, $N = 100$, and $\alpha_0 = 0.2$) the GMM median absolute error is always the smallest. The ranking is less obvious between EL and LIML for this comparison criterion as well. In Table 1 for $T^0 = 10, 50$ and $\alpha_0 = 0.2, 0.8$ EL median absolute error is smaller than LIML. In Table 2 it is smaller for $T^0 = 10$ and $\alpha_0 = 0.2$, however as T gets larger and closer to N, LIML median absolute error becomes smaller, although the difference between them is very small especially when α_0 is large.

5 Conclusion

In this paper we show that in autoregressive panel data models, the EL estimator that uses all the moment conditions based on all the available lags at each period are consistent and asymptotically normal when both N and T tend to infinity. When showing normality, we applied Newey & Windmeijer (2007) method that they use for a general cross sectional model. For the EL estimator, for normality, the required condition on the relative rates at which N and T can grow turns out to be much more strict than that of the LIML and GMM estimators. Under this restriction, the LIML and GMM asymptotic biases disappear. Therefore, all three estimators that we consider have the same asymptotic distribution. To be able to distinguish the finite sample performances of these estimators we consider a Monte Carlo study. This study reveals that GMM always has the largest median bias, although it has the smallest dispersion. However, the ranking between EL and LIML is not obvious when median bias, interquartile range, and mesian absolute error are concerned.

In future work, we plan to extend the current results by looking at the second order asymptotics of the EL estimator. This helps us to understand the nature of the asymptotic bias of EL and hence ables us to compare the three estimators analytically. As a second extension, we plan to relax the assumptions on the initial conditions and the homoscedasticity and study the properties of “robust” LIML estimator.

Appendix

Through out the Appendix, we let C to denote a generic positive constant that may be different in different uses. Also, let $w.p.a.1$ stand for "with probability approaching one". To show the order of each term, we frequently employ results in Alvarez & Arellano (2003). AA refers to the paper of Alvarez & Arellano (2003); also AA (.) reads "the formula (.) in Alvarez & Arellano (2003)".

Lemma 1. $\|\frac{1}{N}\sum_{i=1}^{N} Z_i' v_i^*\| = O_p\left(\frac{T}{\sqrt{N}}\right)$.

Proof. By independence across individuals we have

$$
\begin{aligned}
E\left(\|\frac{1}{N}\sum_{i=1}^{N} Z_i' v_i^*\|^2\right) &= E\left(\frac{1}{N^2}\sum_{i=1}^{N} v_i^{*\prime} Z_i \sum_{i=1}^{N} Z_i' v_i^*\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} E\left(v_i^{*\prime} Z_i Z_i' v_i^*\right) + \frac{2}{N^2}\sum_{i=1}^{N}\sum_{j>i} E\left(v_i^{*\prime} Z_i\right) E\left(Z_j' v_j^*\right) \\
&= \frac{\sigma^2}{N} tr\left\{E\left(Z_i' Z_i\right)\right\} \\
&= \frac{\sigma^2}{N}\sum_{t=1}^{T-1}\sum_{s=0}^{t-1} E\left(y_{is}^2\right). \qquad \text{(A-1)}
\end{aligned}
$$

Recall

$$
\begin{aligned}
E\left(Z_i' Z_i\right) &= E\left\{ \underbrace{\begin{bmatrix} z_{i1} & 0 & \ldots & 0 \\ 0 & z_{i2} & 0\ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & 0 & z_{i,T-1} \end{bmatrix}}_{q\times(T-1)} \underbrace{\begin{bmatrix} z_{i1}' & 0 & \ldots & 0 \\ 0 & z_{i2}' & 0\ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & 0 & z_{i,T-1}' \end{bmatrix}}_{(T-1)\times q} \right\} \\
&= E\begin{bmatrix} z_{i1}z_{i1}' & 0 & \ldots & 0 \\ 0 & z_{i2}z_{i2}' & 0\ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & 0 & z_{i,T-1}z_{i,T-1}' \end{bmatrix},
\end{aligned}
$$

where

$$z_{it}z_{it}' = \begin{bmatrix} y_{i0}^2 & \cdots & y_{i0}y_{i,T-1} \\ y_{i1}y_{i0} & \cdots & y_{i1}y_{i,T-1} \\ \vdots & \ddots & \vdots \\ y_{i,T-1}y_{i0} & \cdots & y_{i,T-1}^2 \end{bmatrix}.$$

Let $\omega_{it} = y_{it} - \frac{\eta_i}{(1-\alpha_0)}$. Then we have

$$\sum_{s=0}^{t-1} y_{is}^2 = \sum_{s=1}^{t} \omega_{i,s-1}^2 - 2t\bar{\bar{\omega}}_{i(-1)} \frac{\eta_i}{(1-\alpha_0)} + t\frac{\eta_i^2}{(1-\alpha_0)^2},$$

where $\bar{\bar{\omega}}_{i(-1)} = \frac{1}{t}\sum_{s=1}^{t} \omega_{i,s-1}$.

Under Assumptions 1–3, we have

$$E\left(\omega_{i,s-1}^2\right) = \frac{\sigma^2}{1-\alpha_0^2},$$

$$E\left(\bar{\bar{\omega}}_{i(-1)} \frac{\eta_i}{(1-\alpha_0)}\right) = 0,$$

$$E\left(\frac{\eta_i^2}{(1-\alpha_0)^2}\right) = \frac{\sigma_\eta^2}{(1-\alpha_0)^2}.$$

Therefore,

$$\sum_{s=0}^{t-1} E\left(y_{is}^2\right) = t\left(\frac{\sigma^2}{1-\alpha_0^2} + \frac{\sigma_\eta^2}{(1-\alpha_0)^2}\right),$$

and for $C = \sigma^2\left(\frac{\sigma^2}{1-\alpha_0^2} + \frac{\sigma_\eta^2}{(1-\alpha_0)^2}\right)$, (A-1) becomes

$$\frac{1}{N}\sum_{t=1}^{T-1} Ct = C\frac{(T-1)T}{2N}.$$

Hence the conclusion follows by Markov's Inequality. □

Lemma 2.

$$E\Big[\sup_{\alpha} \|Z_i'(y_i^* - \alpha x_i^*)\|^\gamma\Big] = O(T^{3\gamma-2}).$$

Proof. Consider

$$
\begin{aligned}
\|Z_i'(y_i^* - \alpha x_i^*)\|^\gamma &= \left[\sum_{t=1}^{T-1}(y_{it}^* - \alpha x_{it}^*)^2 \sum_{s=1}^{t} y_{i,s-1}^2\right]^{\frac{\gamma}{2}} \\
&\leq (T-1)^{\frac{\gamma}{2}-1}\sum_{t=1}^{T-1}\left|y_{it}^* - \alpha x_{it}^*\right|^\gamma \Big(\sum_{s=1}^{t} y_{i,s-1}^2\Big)^{\frac{\gamma}{2}} \\
&= (T-1)^{\frac{\gamma}{2}-1}\sum_{t=1}^{T-1}\left|(\alpha_0-\alpha)x_{it}^* - v_{it}^*\right|^\gamma \Big(\sum_{s=1}^{t} y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}.
\end{aligned}
$$

The inequality in the second line is obtained using Loève's c_r inequality. From AA (A43), we can write

$$
x_{it}^* = \psi_t\Big(y_{i,t-1} - \frac{\eta_i}{1-\alpha}\Big) - c_t\widetilde{v}_{itT}, \text{ where}
$$

$$
c_t = \sqrt{\frac{T-t}{T-t+1}}, \tag{A-2}
$$

$$
\psi_t = c_t\Big(1 - \frac{\alpha\phi_{T-t}}{T-t}\Big), \tag{A-3}
$$

$$
\phi_j = \frac{1-\alpha^j}{1-\alpha}, \text{ and} \tag{A-4}
$$

$$
\widetilde{v}_{itT} = \frac{1}{T-t}(\phi_{T-t}v_{it} + \ldots + \phi_1 v_{i,T-1}). \tag{A-5}
$$

Using this expression for x_{it}^*, we have, for $\beta = \alpha - \alpha_0$,

$$
\begin{aligned}
\|Z_i'(y_i^* - \alpha x_i^*)\|^\gamma &= (T-1)^{\frac{\gamma}{2}-1}\sum_{t=1}^{T-1}\Big|v_{it}^* + \beta c_t\widetilde{v}_{itT} - \beta\psi_t\big(y_{i,t-1} - \frac{\eta_i}{1-\alpha}\big)\Big|^\gamma \Big(\sum_{s=1}^{t} y_{i,s-1}^2\Big)^{\frac{\gamma}{2}} \\
&\leq C(T-1)^{\frac{\gamma}{2}-1}\sum_{t=1}^{T-1}\Big[\left|v_{it}^*\right|^\gamma + \left|\beta c_t\widetilde{v}_{itT}\right|^\gamma + \left|\beta\psi_t y_{i,t-1}\right|^\gamma \\
&\quad + \Big|\beta\psi_t\frac{\eta_i}{1-\alpha}\Big|^\gamma\Big]\Big(\sum_{s=1}^{t} y_{i,s-1}^2\Big)^{\frac{\gamma}{2}},
\end{aligned}
$$

where we used the fact that $\left|\sum_{i=1}^m a_i\right|^\gamma \leq c_\gamma\sum_{i=1}^m |a_i|^\gamma$ with c_γ being $m^{\gamma-1}$.

Taking the expectation of $\sup\limits_{\alpha}\|Z_i'(y_i^* - \alpha x_i^*)\|^\gamma$ we get

$$
\begin{aligned}
E\Big[\sup_{\alpha}\|Z_i'(y_i^* - \alpha x_i^*)\|^\gamma\Big] &\leq C(T-1)^{\frac{\gamma}{2}-1}\sum_{t=1}^{T-1}\Bigg\{E\Big[\left|v_{it}^*\right|^\gamma\Big(\sum_{s=1}^{t} y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}\Big] \\
&\quad + \sup_{\alpha}\left|\beta c_t\right|^\gamma E\Big[\left|\widetilde{v}_{itT}\right|^\gamma\Big(\sum_{s=1}^{t} y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}\Big]
\end{aligned} \tag{A-6}
$$

$$+\sup_{\alpha}\big|\beta\psi_t\big|^{\gamma}E\Big[\big|y_{i,t-1}\big|^{\gamma}\Big(\sum_{s=1}^{t}y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}\Big]$$

$$+\sup_{\alpha}\big|\beta\psi_t\big|^{\gamma}E\Big[\big|\frac{\eta_i}{1-\alpha}\big|^{\gamma}\Big(\sum_{s=1}^{t}y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}\Big]\Big\}.$$

Now, we consider the first, second, and fourth elements on the right hand side. Note that v_{it}^* and $\widetilde{v}_{itT}$ includes the present and future legs of the error terms, i.e., $v_{it}, v_{i,t+1}, \ldots, v_{iT}$, hence by Assumption 1 we have

$$E\Big[\big|v_{it}^*\big|^{\gamma}\Big(\sum_{s=1}^{t}y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}\Big]=E\big|v_{it}^*\big|^{\gamma}E\Big(\sum_{s=1}^{t}y_{i,s-1}^2\Big)^{\frac{\gamma}{2}},$$

$$E\Big[\big|\widetilde{v}_{itT}\big|^{\gamma}\Big(\sum_{s=1}^{t}y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}\Big]=E\big|\widetilde{v}_{itT}\big|^{\gamma}E\Big(\sum_{s=1}^{t}y_{i,s-1}^2\Big)^{\frac{\gamma}{2}},$$

$$E\Big[\big|\frac{\eta_i}{1-\alpha}\big|^{\gamma}\Big(\sum_{s=1}^{t}y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}\Big]=E\big|\frac{\eta_i}{1-\alpha}\big|^{\gamma}E\Big(\sum_{s=1}^{t}y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}.$$

Under Assumption 5 $E\big|v_{it}^*\big|^{\gamma}$, $E\big|\widetilde{v}_{itT}\big|^{\gamma}$, and $E\big|\eta_i\big|^{\gamma}$ are constant. Therefore, the order of magnitude of $E\Big[\sup_{\alpha}\|Z_i'(y_i^*-\alpha x_i^*)\|^{\gamma}\Big]$ is determined by the third term in (A-6).

We have

$$\begin{aligned}E\Big[\big|y_{i(t-1)}\big|^{\gamma}\Big(\sum_{s=1}^{t}y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}\Big] &\leq E\Big[\big|y_{i,t-1}\big|^{\gamma}t^{\frac{\gamma}{2}-1}\sum_{s=1}^{t}\big|y_{i,s-1}\big|^{\gamma}\Big]\\ &= t^{\frac{\gamma}{2}-1}\sum_{s=1}^{t}E\Big[\big|y_{i,t-1}y_{i,s-1}\big|^{\gamma}\Big], \qquad \text{(A-7)}\end{aligned}$$

where we have used Loève's c_r inequality.

Solving (1) recursively we obtain

$$y_{i,t-1}=\alpha^{t-1}y_{i0}+\frac{1-\alpha^{t-1}}{1-\alpha}\eta_i+\sum_{k=0}^{t-2}\alpha^k v_{i,t-1-k} \qquad \text{(A-8)}$$

and using (A-8) along with Loève's c_r inequality in (A-7), and under Assumption 2 and Assumption 5 we get

(A-9)

$$t^{\frac{\gamma}{2}-1}\sum_{s=1}^{t}E\big|\alpha^{s+t-2}y_{i0}^2+\frac{1-\alpha^{t-1}}{1-\alpha}\alpha^{s-1}\eta_i y_{i0}+\sum_{k=0}^{t-2}\alpha^k\alpha^{s-1}v_{i,t-1-k}y_{i0}+\frac{1-\alpha^{s-1}}{1-\alpha}\alpha^{t-1}y_{i0}\eta_i$$

$$
+ \frac{1-\alpha^{s-1}}{1-\alpha}\frac{1-\alpha^{t-1}}{1-\alpha}\eta_i^2 + \sum_{k=0}^{t-2}\alpha^k\frac{1-\alpha^{s-1}}{1-\alpha}v_{i,t-1-k}\eta_i + \sum_{k=0}^{s-2}\alpha^k\alpha^{t-1}y_{i0}v_{i,s-1-k}
$$

$$
+ \sum_{k=0}^{s-2}\alpha^k\frac{1-\alpha^{t-1}}{1-\alpha}v_{i,s-1-k}\eta_i + \sum_{j=0}^{t-2}\sum_{k=0}^{s-2}\alpha^{k+j}v_{i,t-1-j}\eta_i v_{i,s-1-k}\big|^\gamma
$$

$$
\le Ct^{\frac{\gamma}{2}-1}\sum_{s=1}^{t}\alpha^{\gamma(s+t-2)}E\big|y_{i0}^2\big|^\gamma + \big|\frac{1-\alpha^{t-1}}{1-\alpha}\alpha^{s-1}\big|^\gamma E\big|\eta_i\big|^\gamma E\big|y_{i0}\big|^\gamma
$$

$$
+ \big|\alpha^{s-1}\big|^\gamma E\big|y_{i0}\big|^\gamma E\big|\sum_{k=0}^{t-2}\alpha^k\alpha^{s-1}v_{i,t-1-k}\big|^\gamma + \big|\frac{1-\alpha^{s-1}}{1-\alpha}\alpha^{t-1}\big|^\gamma E\big|y_{i0}\big|^\gamma E\big|\eta_i\big|^\gamma
$$

$$
+ \big|\frac{1-\alpha^{s-1}}{1-\alpha}\frac{1-\alpha^{t-1}}{1-\alpha}\big|^\gamma E\big|\eta_i^2\big|^\gamma + \big|\frac{1-\alpha^{s-1}}{1-\alpha}\big|^\gamma E\big|\eta_i\big|^\gamma E\big|\sum_{k=0}^{t-2}\alpha^k v_{i,t-1-k}\big|^\gamma
$$

$$
+ \big|\alpha^{t-1}\big|^\gamma E\big|y_{i0}\big|^\gamma E\big|\sum_{k=0}^{s-2}\alpha^k v_{i,s-1-k}\big|^\gamma + \big|\frac{1-\alpha^{t-1}}{1-\alpha}\big|^\gamma E\big|\eta_i\big|^\gamma E\big|\sum_{k=0}^{s-2}\alpha^k v_{i,s-1-k}\big|^\gamma
$$

$$
+ E\big|\sum_{j=0}^{t-2}\sum_{k=0}^{s-2}\alpha^{k+j}v_{i(t-1-j)}\eta_i v_{i,s-1-k}\big|^\gamma
$$

$$
\le Ct^{\frac{\gamma}{2}-1}\Big\{\frac{1-|\alpha^\gamma|^t}{1-|\alpha^\gamma|} + \left|\frac{1-\alpha^{t-1}}{1-\alpha}\right|\frac{\left(1-|\alpha^\gamma|^t\right)}{\left(1-|\alpha^\gamma|\right)} + (t-1)^{\gamma-1}\frac{\left(1-|\alpha^\gamma|^t\right)\left(1-|\alpha^\gamma|^{t-1}\right)}{\left(1-|\alpha^\gamma|\right)^2}
$$

$$
+ |\alpha^\gamma|^{t-1}\frac{1}{|1-\alpha|^\gamma}\left(t+\frac{1-|\alpha^\gamma|^t}{1-|\alpha^\gamma|}\right) + \left|\frac{1-\alpha^{t-1}}{1-\alpha}\right|^\gamma\left|\frac{1}{1-\alpha}\right|^\gamma\left(t+\frac{1-|\alpha^\gamma|^t}{1-|\alpha^\gamma|}\right)
$$

$$
+ \frac{1}{|1-\alpha|^\gamma}\left(t+\frac{1-|\alpha^\gamma|^t}{1-|\alpha^\gamma|}\right)\frac{\left(1-|\alpha^\gamma|^t\right)}{\left(1-|\alpha^\gamma|\right)}(t-1)^{\gamma-1} + |\alpha|^{t-1}\frac{1}{\left(1-|\alpha^\gamma|\right)}\sum_{s=1}^{t}(s-1)^{\gamma-1}
$$

$$
+ \left|\frac{1-\alpha^{t-1}}{1-\alpha}\right|\frac{1}{\left(1-|\alpha^\gamma|\right)}\sum_{s=1}^{t}(s-1)^{\gamma-1} + (t-1)^{\gamma-1}\frac{1}{\left(1-|\alpha^\gamma|\right)^2}\sum_{s=1}^{t-1}s^{\gamma-1}.
$$

Hence, to determine the order of magnitude of $E\Big[\sup_\alpha \|Z_i'(y_i^* - \alpha x_i^*)\|^\gamma\Big]$ we consider

$$
C(T-1)^{\frac{\gamma}{2}-1}\sum_{t=1}^{T-1}\sup_\alpha\big|\beta\psi_t\big|^\gamma E\Big[\big|\omega_{i,t-1}\big|^\gamma\Big(\sum_{s=1}^{t}y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}\Big]
$$

$$
\le C(T-1)^{\frac{\gamma}{2}-1}\sum_{t=1}^{T-1}E\Big[\big|\omega_{i,t-1}\big|^\gamma\Big(\sum_{s=1}^{t}y_{i,s-1}^2\Big)^{\frac{\gamma}{2}}\Big]
$$

$$
\le C(T-1)^{\frac{\gamma}{2}-1}\sum_{t=1}^{T-1}Ct^{\frac{\gamma}{2}-1}\Big\{\frac{1}{1-|\alpha^\gamma|} + \frac{1}{|1-\alpha|\left(1-|\alpha^\gamma|\right)} + (t-1)^{\gamma-1}\frac{1}{\left(1-|\alpha^\gamma|\right)^2}
$$

$$
+ \frac{1}{|1-\alpha|^\gamma}\left(t+\frac{1}{1-|\alpha^\gamma|}\right) + \frac{1}{|1-\alpha|^{2\gamma}}\left(t+\frac{1}{1-|\alpha^\gamma|}\right)
$$

$$+\frac{1}{|1-\alpha|^{\gamma}}\left(t+\frac{1}{1-|\alpha^{\gamma}|}\right)\frac{1}{(1-|\alpha^{\gamma}|)}(t-1)^{\gamma-1}+\frac{1}{(1-|\alpha^{\gamma}|)}\sum_{s=1}^{t}(s-1)^{\gamma-1}$$
$$+\frac{1}{|1-\alpha|\,(1-|\alpha^{\gamma}|)}\sum_{s=1}^{t}(s-1)^{\gamma-1}+(t-1)^{\gamma-1}\frac{1}{(1-|\alpha^{\gamma}|)^{2}}\sum_{s=1}^{t-1}s^{\gamma-1}\Big\}$$
$$=O(T^{3\gamma-2}),$$

where the first inequality uses the fact that $\left|\beta\psi_t\right|^{\gamma}$ is bounded in t and α and the second inequality uses the fact that $1-|\alpha^{\gamma}|^{t}$, $1-\alpha^{t-1}$ and $|\alpha|^{t-1}$ are bounded in t. For the last result we used the fact that $T^{-(v+1)}\sum_{t=1}^{T}t^{v}\rightarrow 1/(v+1)$ (cf. Hamilton (1994), Proposition 17.4 (h)). □

Proof of Theorem 1. Combining the result in Lemma 2 with the first result in Appendix of Guggenberger & Smith (2005), we have

(A-10) $\max_{i\le N}\sup_{\alpha}\|Z_i'(y_i^*-\alpha x_i^*)\|=O_p(N^{1/\gamma}(E[\sup_{\alpha}\|Z_i'(y_i^*-\alpha x_i^*)\|^{\gamma}])^{1/\gamma})=O_p(N^{1/\gamma}T^{3-2/\gamma}).$

By the hypothesis of the theorem, namely $N^{1/\gamma}T^{3-2/\gamma}\sqrt{T^2/N}\rightarrow 0$ for $\gamma>2$, there exists τ_N such that $\frac{T}{\sqrt{N}}=o(\tau_N)$ and
$\tau_N=o(N^{-1/\gamma}T^{-3+2/\gamma})$. Let $L_N=\{\lambda:\|\lambda\|\le\tau_N\}$. Note that

$$\sup_{\lambda\in L_N,|\alpha|<1,i\le N}\left|\lambda' Z_i'(y_i^*-\alpha x_i^*)\right|\le\tau_N\max_{i\le N}\sup_{|\alpha|<1}\left\|Z_i'(y_i^*-\alpha x_i^*)\right\|=O_p(\tau_N N^{1/\gamma}T^{3-2/\gamma})\rightarrow 0.$$

Note that the multipliers of the moment conditions, λ, have to satisfy the condition $\lambda' Z_i'(y_i^*-\alpha x_i^*)>-1$, for all $i=1,\ldots,N$. Let $\widehat{L}(\alpha)$ be the set of λs that satisfies this condition, i.e. $\widehat{L}(\alpha)=\{\lambda:\lambda' Z_i'(y_i^*-\alpha x_i^*)>-1,i=1,\ldots,N\}$. Therefore, there exists a C such that $w.p.a.1$, for all $|\alpha|\le 1$, $\lambda\in L_N$, $i\le N$

(A-11) $L_N\subset\widehat{L}(\alpha),\ -C\le\dfrac{-1}{[1+\lambda' Z_i'(y_i^*-\alpha x_i^*)]^2}\le -C^{-1},\ \left|\dfrac{1}{[1+\lambda' Z_i'(y_i^*-\alpha x_i^*)]^3}\right|\le C.$

Let $\widehat{P}(\alpha,\lambda)=\frac{1}{NT}\sum_{i=1}^{N}\ln\left[1+\lambda' Z_i'(y_i^*-\alpha x_i^*)\right]$. By a Taylor expansion around $\lambda=0$ with Lagrange remainder for all $\lambda\in L_N$

$$\widehat{P}(\alpha,\lambda)=\lambda'\frac{1}{NT}\sum_{i=1}^{N}Z_i'(y_i^*-\alpha x_i^*)-\lambda'\frac{1}{T}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{Z_i'(y_i^*-\alpha x_i^*)(y_i^*-\alpha x_i^*)'Z_i}{\left[1+\overline{\lambda}' Z_i'(y_i^*-\alpha x_i^*)\right]^2}\right]\lambda$$

where $\bar{\lambda}$ lies between $\widehat{\lambda}$ and 0. By Assumption 4, $Lemma A0$ of Newey & Windmeijer (2007), we have, $w.p.a.1$ $\lambda_{\min}\big(\frac{1}{N}\sum_{i=1}^{N} Z_i'(y_i^*-\alpha x_i^*)(y_i^*-\alpha x_i^*)'Z_i\big) \geq C^{-1}$ and $\lambda_{\max}\big(\frac{1}{N}\sum_{i=1}^{N} Z_i'(y_i^*-\alpha x_i^*)(y_i^*-\alpha x_i^*)'Z_i\big) \geq C$. Then, $w.p.a.1$ for all $|\alpha|<1$ and $\lambda \in L_N$,

(A-12)

$$\lambda'\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^*-\alpha x_i^*) - C\frac{1}{T}\|\lambda\|^2 \leq \widehat{P}(\alpha,\lambda) \leq \lambda'\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^*-\alpha x_i^*) - C^{-1}\frac{1}{T}\|\lambda\|^2$$
$$\leq \|\lambda\|\|\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^*-\alpha x_i^*)\| - C^{-1}\frac{1}{T}\|\lambda\|^2$$

Let $\widetilde{\lambda} = \arg\max_{\lambda\in L_N} \widehat{P}(\alpha_0,\lambda)$. By the right hand inequality in eq. (A-12),

$$0 = \widehat{P}(\alpha_0,\lambda) \leq \widehat{P}(\alpha_0,\widetilde{\lambda}) \leq \|\widetilde{\lambda}\|\|\frac{1}{NT}\sum_{i=1}^{N} Z_i'v_i^*\| - C^{-1}\frac{1}{T}\|\widetilde{\lambda}\|^2.$$

Subtracting $-C^{-1}\frac{1}{T}\|\widetilde{\lambda}\|^2$ from both sides and dividing through by $-C^{-1}\frac{1}{T}\|\widetilde{\lambda}\|$ and using the result of Lemma 1 gives

$$\|\widetilde{\lambda}\| \leq C\|\frac{1}{N}\sum_{i=1}^{N} Z_i'v_i^*\| = O_p(\frac{T}{\sqrt{N}}).$$

Now, following the same arguments in the proof of Lemma A3 of Newey & Windmeijer (2007), it can be shown that $\|\widehat{\lambda}\| = O_p(\frac{T}{\sqrt{N}})$.

Now, expanding around $\lambda = 0$ (note that we let $\widehat{\alpha} \equiv \widehat{\alpha}_{EL}$ for simplification) to obtain

$$\begin{aligned}
\widehat{Q}(\widehat{\alpha}) = \widehat{P}(\widehat{\alpha},\widehat{\lambda}) &= \frac{1}{NT}\sum_{i=1}^{N}\ln 1 + \frac{1}{NT}\sum_{i=1}^{N}\frac{(y_i^*-\widehat{\alpha}x_i^*)'Z_i}{[1+\lambda'Z_i'(y_i^*-\widehat{\alpha}x_i^*)]}|_{\lambda=0}\widehat{\lambda} \\
&- \frac{1}{2}\frac{1}{NT}\widehat{\lambda}'\left[\sum_{i=1}^{N}\frac{Z_i'(y_i^*-\widehat{\alpha}x_i^*)(y_i^*-\widehat{\alpha}x_i^*)'Z_i}{[1+\lambda'Z_i'(y_i^*-\widehat{\alpha}x_i^*)]^2}|_{\lambda=0}\right]\widehat{\lambda} \\
&+ \frac{1}{3}\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\left[1+\bar{\lambda}'Z_i'(y_i^*-\widehat{\alpha}x_i^*)\right]^3}\left[(y_i^*-\widehat{\alpha}x_i^*)'Z_i\hat{\lambda}\right]^3 \\
&= \frac{1}{NT}\sum_{i=1}^{N}(y_i^*-\widehat{\alpha}x_i^*)'Z_i\widehat{\lambda} - \frac{1}{2}\frac{1}{NT}\widehat{\lambda}'\sum_{i=1}^{N}Z_i'(y_i^*-\widehat{\alpha}x_i^*)(y_i^*-\widehat{\alpha}x_i^*)'Z_i\widehat{\lambda} \\
&+ \hat{r}
\end{aligned}$$

where

$$\hat{r} = \frac{1}{3}\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\left[1+\bar{\lambda}'Z_i'(y_i^*-\widehat{\alpha}x_i^*)\right]^3}\left[(y_i^*-\widehat{\alpha}x_i^*)'Z_i\hat{\lambda}\right]^3.$$

and $\|\bar{\lambda}\| \leq \|\widehat{\lambda}\|$. We have, $w.p.a.1$

$$\begin{aligned}
|\widehat{r}| &\leq \frac{1}{3}\|\widehat{\lambda}\| \max_{1\leq i\leq N}\sup_{\alpha}\|Z_i'(y_i^* - \widehat{\alpha}x_i^*)\|\frac{1}{T}\widehat{\lambda}'\left[\frac{1}{N}\sum_{i=1}^{N}\frac{Z_i'(y_i^* - \widehat{\alpha}x_i^*)(y_i^* - \widehat{\alpha}x_i^*)'Z_i}{\left[1+\bar{\lambda}'Z_i'(y_i^* - \widehat{\alpha}x_i^*)\right]^3}\right]\widehat{\lambda} \\
&\leq C\|\widehat{\lambda}\| \max_{1\leq i\leq N}\sup_{\alpha}\|Z_i'(y_i^* - \widehat{\alpha}x_i^*)\|\frac{1}{T}\widehat{\lambda}'\left[\frac{1}{N}\sum_{i=1}^{N}Z_i'(y_i^* - \widehat{\alpha}x_i^*)(y_i^* - \widehat{\alpha}x_i^*)'Z_i\right]\widehat{\lambda} \\
&\leq O_p(\sqrt{T^2/N}N^{1/\gamma}T^{3-2/\gamma})C\frac{1}{T}\|\widehat{\lambda}\|^2 = o_p(\frac{T}{N}).
\end{aligned}$$

Also, $\widehat{\lambda}$ solves the equations:

$$\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{[1+\lambda'Z_i'(y_i^* - \widehat{\alpha}x_i^*)]}Z_i'(y_i^* - \widehat{\alpha}x_i^*) = 0.$$

Expanding around $\lambda = 0$:

$$\begin{aligned}
0 &= \frac{1}{NT}\sum_{i=1}^{N}\frac{1}{[1+\lambda'Z_i'(y_i^* - \widehat{\alpha}x_i^*)]}Z_i'(y_i^* - \widehat{\alpha}x_i^*)\mid_{\lambda=0} \\
&- \frac{1}{NT}\sum_{i=1}^{N}\frac{1}{[1+\lambda'Z_i'(y_i^* - \widehat{\alpha}x_i^*)]^2}Z_i'(y_i^* - \widehat{\alpha}x_i^*)(y_i^* - \widehat{\alpha}x_i^*)'Z_i\mid_{\lambda=0}\widehat{\lambda} \\
&+ \frac{1}{2}\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\left[1+\bar{\lambda}'Z_i'(y_i^* - \widehat{\alpha}x_i^*)\right]^3}\left[(y_i^* - \widehat{\alpha}x_i^*)'Z_i\hat{\lambda}\right]^2 Z_i'(y_i^* - \widehat{\alpha}x_i^*)
\end{aligned}$$

$$\Rightarrow 0 = \frac{1}{NT}\sum_{i=1}^{N}Z_i'(y_i^* - \widehat{\alpha}x_i^*) - \frac{1}{NT}\sum_{i=1}^{N}Z_i'(y_i^* - \widehat{\alpha}x_i^*)(y_i^* - \widehat{\alpha}x_i^*)'Z_i\widehat{\lambda} + \widehat{R}$$

where

$$\widehat{R} = \frac{1}{2}\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\left[1+\bar{\lambda}'Z_i'(y_i^* - \widehat{\alpha}x_i^*)\right]^3}\left[(y_i^* - \widehat{\alpha}x_i^*)'Z_i\hat{\lambda}\right]^2 Z_i'(y_i^* - \widehat{\alpha}x_i^*)$$

and $\bar{\lambda}$ lies in between 0 and $\widehat{\lambda}$. Note that $\max_{i\leq N}|\bar{\lambda}'Z_i'(y_i^* - \widehat{\alpha}x_i^*)| \leq \max_{i\leq N}|\widehat{\lambda}'Z_i'(y_i^* - \widehat{\alpha}x_i^*)| \leq \|\widehat{\lambda}\| \max_{1\leq i\leq N}\sup_{\alpha}\|Z_i'(y_i^* - \alpha x_i^*)\| \to 0$.

Then we have

$$\begin{aligned}
\|\widehat{R}\| &\leq C\max_{1\leq i\leq N}\Big(\sup_{\alpha}\|Z_i'(y_i^* - \alpha x_i^*)\|\frac{1}{\left|1+\bar{\lambda}'Z_i'(y_i^* - \widehat{\alpha}x_i^*)\right|^3}\Big) \\
&\times \frac{1}{T}\widehat{\lambda}'\Big(\frac{1}{N}\sum_{i=1}^{N}Z_i'(y_i^* - \alpha x_i^*)(y_i^* - \alpha x_i^*)'Z_i\Big)\widehat{\lambda} \\
&\leq C\max_{1\leq i\leq N}\sup_{\alpha}\|Z_i'(y_i^* - \alpha x_i^*)\|\frac{1}{T}\|\widehat{\lambda}\|^2 = O_p(N^{1/\gamma}T^{3-2/\gamma}T/N) = o_p(\sqrt{T/N}).
\end{aligned}$$

Solving for $\widehat{\lambda}$:

$$\widehat{\lambda} = \left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)(y_i^* - \widehat{\alpha}x_i^*)'Z_i\right]^{-1}\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*) + \widehat{R}\right].$$

Plugging into $\widehat{Q}(\widehat{\alpha})$:

(A-13)

$$\begin{aligned}
\widehat{Q}(\widehat{\alpha}) =& \left[\frac{1}{NT}\sum_{i=1}^{N}(y_i^* - \widehat{\alpha}x_i^*)'Z_i\right]\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)(y_i^* - \widehat{\alpha}x_i^*)'Z_i\right]^{-1} \\
&\times\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*) + \widehat{R}\right] - \frac{1}{2}\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*) + \widehat{R}\right]' \\
&\times\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)(y_i^* - \widehat{\alpha}x_i^*)'Z_i\right]^{-1}\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*) + \widehat{R}\right] + o_p(\frac{T}{N}) \\
=& \left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)\right]'\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)(y_i^* - \widehat{\alpha}x_i^*)'Z_i\right]^{-1} \\
&\times\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)\right] - \frac{1}{2}\widehat{R}'\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)(y_i^* - \widehat{\alpha}x_i^*)'Z_i\right]^{-1}\widehat{R} \\
&+ o_p(\frac{T}{N}) \\
\leq& \left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)\right]'\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)(y_i^* - \widehat{\alpha}x_i^*)'Z_i\right]^{-1} \\
&\times\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)\right] - \|\widehat{R}\|^2 C + o_p(\frac{T}{N}) \\
=& \left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)\right]'\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)(y_i^* - \widehat{\alpha}x_i^*)'Z_i\right]^{-1} \\
&\times\left[\frac{1}{NT}\sum_{i=1}^{N} Z_i'(y_i^* - \widehat{\alpha}x_i^*)\right] + o_p(\frac{T}{N}).
\end{aligned}$$

The first term on the right side of (A-13) is the objective function of continuously updating GMM estimator (CUE). Hence the last conclusion shows that the difference of the CUE and EL objective functions converges uniformly to zero in α. The remainder of the proof then follows from the proof for of Theorem 3 (the consistency of the LIML estimator) of AA upon noting that in a linear model under homoskedasticity CUE the LIML estimator as mentioned in section 2.2. □

Lemma 3. *If Assumption 4 (i) holds, Assumption 8 (ii) of Newey & Windmeijer (2007) also holds.*

Proof. In Newey & Windmeijer (2007) notation

$$
\begin{aligned}
&|a\big[\Omega^k(\widetilde{\alpha}) - \Omega^k(\alpha)\big]b| = |a\big[E[Z_i'(y_i^* - \widetilde{\alpha}x_i^*)(-x_i^*)'Z_i] - E[Z_i'(y_i^* - \alpha x_i^*)(-x_i^*)'Z_i]\big]b| \\
&= |a\big[E[Z_i'W_i^*\begin{pmatrix} 1 \\ -\widetilde{\alpha} \end{pmatrix}\begin{pmatrix} 0 & -1 \end{pmatrix}W_i^{*\prime}Z_i] - E[Z_i'W_i^*\begin{pmatrix} 1 \\ -\alpha \end{pmatrix}\begin{pmatrix} 0 & -1 \end{pmatrix}W_i^{*\prime}Z_i]\big]b| \\
&= |a\big[E[Z_i'W_i^*\begin{pmatrix} 0 & 1 \\ 0 & \widetilde{\alpha} \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ 0 & \alpha \end{pmatrix}W_i^{*\prime}Z_i]\big]b| \\
&= |tr\Big\{\begin{pmatrix} 0 & 0 \\ 0 & \widetilde{\alpha} - \alpha \end{pmatrix}E\big[W_i^{*\prime}Z_iba'Z_i'W_i^*\big]\Big\}| \\
&= |\widetilde{\alpha} - \alpha||E[(x_i^{*\prime}Z_iba'Z_i'x_i^*)]| \\
&= |\widetilde{\alpha} - \alpha||E[(a'Z_i'x_i^*)(b'Z_i'x_i^*)]| \\
&\leq |\widetilde{\alpha} - \alpha|E[(a'Z_i'x_i^*)^2]^{1/2}E[(b'Z_i'x_i^*)^2]^{1/2} \\
&\leq C|\widetilde{\alpha} - \alpha|\|a\|\|b\|,
\end{aligned}
$$

where for the second to last inequality, we used Cauchy-Schwartz inequality and the last inequality is obtained by noting that

$$E[(a'Z_i'x_i^*)^2] = a'E[Z_i'x_i^*x_i^{*\prime}Z_i]a \leq \|a\|^2\lambda_{\max}\{E[Z_i'x_i^*x_i^{*\prime}Z_i]\} \leq C\|a\|^2,$$

by Rayleigh quotient and Assumption 4.

The other parts of 8 (ii) follow by $\Omega^{k,l}(\alpha)$ and $\Omega^{kl}(\alpha)$ not depending on α. □

Proof of Theorem 2. We proceed by verifying all the hypotheses of Theorem 3 of Newey & Windmeijer (2007). Note that $Z_i'(y_i^* - \alpha x_i^*)$ is twice continuously differentiable and that its first derivative does not depend on α, so Assumption 7 is satisfied. Also, for our case, Assumption 9 (i) of Newey & Windmeijer (2007) holds from the hypothesis of the theorem, namely $N^{1/\gamma}T^{3-2/\gamma}\sqrt{T^2/N} \to 0$ for $\gamma > 2$.

Now, we show that their Assumption 6 and Assumption 9 (ii) also hold. In Newey & Windmeijer (2007) notation, we have

$$\Big(E[\|g_i\|^4] + E[\|G_i\|^4]\Big)\frac{m}{n} = \Big(E[\|Z_i'v_i^*\|^4] + E[\|Z_i'x_i^*\|^4]\Big)\frac{T}{N}$$
$$\leq \Big(CE[\|Z_i\|^4] + E[\|x_i^*\|^4\|Z_i\|^4]\Big)\frac{T}{N}.$$

The order of magnitude of the second term in the summation dominates that of the first term. Therefore, it is sufficient to show that $\Big(E[\|Z_i'x_i^*\|^4]\Big)\frac{T}{N} \to 0$.

Following the similar steps as in Lemma 2, we note that $E[\|Z_i'x_i^*\|^\gamma] = E[\|Z_i'(y_i^* - \alpha x_i^*)\|^\gamma] = O(T^{3\gamma-2})$. Hence, for $\gamma = 4$, we have

$$E[\|Z_i'x_i^*\|^4] = O(T^{10})$$

Hence, the first part of Assumption 6 and Assumption 9 (ii) of Newey & Windmeijer (2007) hold since $T^{11}/N \to 0$.

The second part of their Assumption 6 follows from Assumption 4 (i), and the rest holds by the model being linear in α.

The parts of Assumption 8 (i) of Newey & Windmeijer (2007) follow similarly upon noting that, for $W_i^* = (y_i^* : x_i^*)$ by Assumption 4 (i) we have

$$\sup_\alpha \Big\| \frac{1}{N}\sum_{i=1}^N Z_i'(y_i^* - \alpha x_i^*)(y_i^* - \alpha x_i^*)'Z_i - E[Z_i'(y_i^* - \alpha x_i^*)(y_i^* - \alpha x_i^*)'Z_i]\Big\|$$
$$= \sup_\alpha \Big\| \begin{pmatrix} 1 & -\alpha \end{pmatrix} \left(\frac{1}{N}\sum_{i=1}^N W_i^{*\prime}Z_iZ_i'W_i^* - E[W_i^{*\prime}Z_iZ_i'W_i^*]\right) \begin{pmatrix} 1 \\ -\alpha \end{pmatrix} \Big\| \to 0.$$

Hence, using Newey & Windmeijer (2007) notation, we have

$$\widehat{\Omega}^k(\alpha) - \Omega^k(\alpha) = \frac{1}{N}\sum_{i=1}^{N} Z_i'(y_i^* - \alpha x_i^*)(-x_i^*)'Z_i - E[Z_i'(y_i^* - \alpha x_i^*)(-x_i^*)'Z_i]$$

$$= \begin{pmatrix} 1 & -\alpha \end{pmatrix} \left(\frac{1}{N}\sum_{i=1}^{N} W_i^{*\prime} Z_i Z_i' W_i^* - E[W_i^{*\prime} Z_i Z_i' W_i^*] \right) \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

$$\widehat{\Omega}^{k,l}(\alpha) - \Omega^{k,l}(\alpha) = \frac{1}{N}\sum_{i=1}^{N} Z_i'(x_i^*)(x_i^*)'Z_i - E[Z_i'(x_i^*)(x_i^*)'Z_i]$$

$$= \begin{pmatrix} 0 & -1 \end{pmatrix} \left(\frac{1}{N}\sum_{i=1}^{N} W_i^{*\prime} Z_i Z_i' W_i^* - E[W_i^{*\prime} Z_i Z_i' W_i^*] \right) \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \text{and}$$

$$\widehat{\Omega}^{kl}(\alpha) = \Omega^{kl}(\alpha) = 0.$$

Finally, assumption 8 (ii) is satisfied by Lemma 3. □

References

Ahn, S. C. & Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal Econometrics*, *68*, 5–27.

Alonso-Borrego, C. & Arellano, M. (1999). Systematically normalized instrumental-variable estimation using panel data. *Journal of Business & Econometric Statistics*, *17*, 36–49.

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

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

Arellano, M. (2003). *Panel Data Econometrics.* Oxford University Press, New York, NY, U.S.A.

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

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, *62*, 657–681.

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

Bond, S. R. (2002). Dynamic panel data models: a guide to micro data method and practice. *Portuguese Economic Journal* , *1*, 141–162.

Bowsher, C. G. (2002). On testing overidentifying restrictions in dynamic panel data models. *Economics Letters*, *77*, 211–220.

Brown, B. & Newey, W. (2001). GMM, efficient bootstrapping, and improved inference. *Journal of Business & Econometric Statistics*, *20*, 507–517.

Gonzalez, A. (2007). Empirical likelihood: improved inference within dynamic panel data models. ESE Discussion Papers 154, Edinburgh School of Economics, University of Edinburgh.

Guggenberger, P. & Smith, R. J. (2005). Generalized empirical likelihood estimators and test under partial, weak, and strong identification. *Econometric Theory*, *21*, 667–709.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press, Princeton, NJ, U.S.A.

Han, C. & Phillips, P C. B. (2006). GMM with many moment conditions. *Econometrica*, *74*, 147–192.

Hansen, L. P., Heaton, J., & Yaron, A. (1996). Finite sample properties of some alternative GMM estimators obtained from financial market data. *Journal of Business & Econometric Statistics*, *14*, 262–280.

Holtz-Eakin, D., Newey, W., & Rosen, H. (1988). Estimating vector autoregressions with panel data. *Econometrica*, *56*, 1371–1395.

Imbens, G. W. (1997). One step estimators for over-identified generalized method of moments models. *Review of Economics Studies*, *64*, 359–383.

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics*, *68*, 53–78.

Kunitomo, N. (1980). Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations. *Journal of the American Statistical Association*, *75*, 693–700.

Morimune, K. (1983). Approximate distribution of k-class estimators when the degree of overidentifiability is large compared with the sample size. *Econometrica*, *51*, 821–841.

Newey, W. & Smith, R. J. (2004). Higher-order properties of GMM and generalized empirical likelihood estiamtors. *Econometrica*, *72*, 219–255.

Newey, W. & Windmeijer, F. (2007). GMM with many weak moment conditions. Working paper, MIT.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*, 1417–1426.

Oĝuzoĝlu, U. (2006). Empirical likelihood estimation of dynamic panel data models with fixed effects. Unpublished manuscript.

Qin, J. & Lawless, J. (1994). Empirical likelihood and generalized estimating equations. *Annals of Statistics*, *22*, 300–325.

Sevestre, P. & Trognon, A. (1985). A note on autoregressive error components models. *Journal of Econometrics*, *28*, 231–245.

Zilliak, J. P. (1997). Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators. *Journal of Business & Econometric Statistics*, *15*, 419–431.