

2002

# 2002-10 Adaptive Estimation of the Dynamic Linear Model with Fixed Effects

Tiemen Woutersen

Marcel Voia

Follow this and additional works at: <https://ir.lib.uwo.ca/economicsresrpt>



Part of the [Economics Commons](#)

---

## Citation of this paper:

Woutersen, Tiemen, Marcel Voia. "2002-10 Adaptive Estimation of the Dynamic Linear Model with Fixed Effects." Department of Economics Research Reports, 2002-10. London, ON: Department of Economics, University of Western Ontario (2002).

**RESEARCH REPORT 2002-10**

**Adaptive Estimation of the Dynamic Linear Model  
with Fixed Effects**

by

**Tiemen Woutersen and Marcel Voia**

September 2002

Department of Economics  
Social Science Centre  
University of Western Ontario  
London, Ontario, Canada  
N6A 5C2  
[econref@uwo.ca](mailto:econref@uwo.ca)

# Adaptive Estimation of the Dynamic Linear Model with Fixed Effects\*

By Tiemen Woutersen and Marcel Voia<sup>†</sup>

September 2002

## Abstract

This paper shows how the dynamic linear model with fixed regressors can be efficiently estimated. This dynamic model can be used to distinguish spurious correlation from state dependence and we show that the integrated likelihood estimator is adaptive for any asymptotics with  $T$  increasing where  $T$  is the number of observations per individual.

KEYWORDS: Panel data, Efficient Estimation, Bayesian Analysis

JEL CLASSIFICATION: C31, C33, C11, C14

## 1 Introduction

THE ANALYSIS OF THE DYNAMIC LINEAR MODEL with fixed effects has been subject of some attention in econometrics for slightly more than two decades, starting with Nickell (1981) and Anderson and Hsiao (1982). The popularity of this linear model might be due to the fact that it is the simplest model in which Heckman's (1981a and 1981b) spurious correlation and state dependence can be studied. Another reason for the popularity of the dynamic linear model is that it can be used for studying the dynamic version of Solow's (1956) growth model, see for example Mankiw, Romer and Weil (1992).

Nickell (1981) shows that the maximum likelihood estimation of the dynamic linear model with fixed effects suffers from the incidental parameter problem of Neyman and Scott (1948).

---

\*We gratefully acknowledge stimulating suggestions from Manuel Arellano, Cheng Hsiao and Tony Lancaster. The CIBC Human Capital and Productivity Program provided financial support. All errors are ours.

<sup>†</sup>Correspondence addresses: University of Western Ontario, Department of Economics, Social Science Centre, London, Ontario, N6A 5C2, Canada. Email: twouters@uwo.ca and mvoia@uwo.ca.

In particular, Nickell shows that the inconsistency of the maximum likelihood estimator is  $O(T^{-1})$  where  $T$  is the number of observations per individual. Econometricians have subsequently developed moment estimators. Examples include Anderson and Hsiao (1982), Holtz-Eakin, Newey and Rosen (1988), Arellano and Bond (1991), Ahn and Schmidt (1995), Blundell and Bond (1998). Combining moment restrictions can be difficult, especially, when some moments become uninformative for particular regions of the parameter space. Newey and Smith (2001) give a general discussion on how combining moments usually causes a higher order bias.

Hsiao et al (2002) react to the problem of combining moments by deriving an estimator based on the likelihood. This paper also uses the likelihood as a starting point but the results do not rely on normality of the error term. Moreover, the distribution of  $y_0$ , the dependent variable in period zero, is left unrestricted.

Lancaster (2002) proposes to approximately separate the parameters of interest from the nuisance parameters. In particular, Lancaster uses a parametrization of the likelihood that has a block diagonal information matrix. That is, the cross derivatives of the log likelihood of the nuisance parameters and parameters of interest is zero in expectation. Lancaster (2002) then integrates out all fixed effects. Woutersen (2001) shows that information-orthogonality reduces the bias of this integrated likelihood estimator to  $O(T^{-2})$  and that the integrated likelihood estimator is asymptotically unbiased and adaptive if  $T \propto N^\alpha$  where  $\alpha > \frac{1}{3}$  and  $N$  is the number of individuals. That is, the integrated likelihood estimator is as efficient as the infeasible maximum likelihood estimator that assumes the values of the nuisance parameters to be known.

Alvarez and Arellano (1998) develop an alternative asymptotic where the number of individuals,  $N$ , increases as well as the number of observations per individual,  $T$ . Using this alternative asymptotics, Hahn and Kuersteiner (2002) develop an expression for the bias for the case in which  $T \propto N$ . Then, Hahn and Kuersteiner (2002) developed an estimator for the dynamic linear model without regressors that is efficient as long as  $T$  increasing as fast as  $N$ . Hahn, Hausman and Kuersteiner (2001) also developed a bias corrected maximum

likelihood estimator but use long differences. Their bias-corrected GMM and Nagar-type estimator reaches the efficiency bound in the same asymptotics as the estimator of Hahn and Kuersteiner (2002).

In most panel datasets,  $T$  is much smaller than  $N$ , as is discussed in the overviews by Chamberlain (1984), Hsiao (1986) and Baltagi (1995). It is therefore desirable to have an efficiency or adaptiveness result that allows  $T$  to increase at an arbitrarily slow rate instead of requiring that  $T \propto N$ . This paper derives such an adaptiveness result without strengthening the conditions of Hahn and Kuersteiner (2002) or Hahn, Hausman and Kuersteiner (2001). In particular, we show that the integrated likelihood estimator reaches the efficiency bound for  $T$  increasing arbitrarily slowly while  $N$  can be fixed or increasing. Interestingly, the efficiency result is an adaptiveness result if Lancaster's (2002) parametrization with a block diagonal information matrix is used. That is, the asymptotic variance of the integrated likelihood estimator is the same as asymptotic variance of an infeasible maximum likelihood estimator that uses the true value of a reparameterized fixed effect.

This paper is organized as follows. Section 2 reviews the integrated likelihood estimator and information orthogonality, and section 3 applies these to the dynamic linear model. Section 4 gives adaptiveness results for the integrated likelihood estimator. Section 5 gives simulation results and section 6 concludes.

## 2 The Integrated Likelihood Estimator and Orthogonality

Suppose we observe  $N$  individuals for  $T$  periods. Let the log likelihood contribution of the  $t^{\text{th}}$  spell of individual  $i$  be denoted by  $L^{it}$ . Summing over the contribution of individual  $i$  yields the log likelihood contribution,

$$L^i(\rho, \lambda_i) = \sum_t L^{it}(\rho, \lambda_i),$$

where  $\rho$  is the common parameter and  $\lambda_i$  is the individual specific effect. Suppose that the parameter  $\rho$  is of interest and that the fixed effect  $\lambda_i$  is a nuisance parameter that controls for heterogeneity. This paper considers elimination of nuisance parameters by integration. This Bayesian treatment of nuisance parameters is straightforward: Formulate a prior on all the

nuisance parameters and then integrate the likelihood with respect to that prior distribution of the nuisance parameters, see Gelman et al. (1995) for an overview. For a panel data model with fixed effects this means that we have to specify priors on the common parameters and all the fixed effects. Berger et al. (1999) review integrated likelihood methods in which flat priors are used for both the parameter of interest and the nuisance parameters. The individual specific nuisance parameters are then eliminated by integration. We denote the logarithm of the integrated likelihood contribution by  $L^{i,I}$ , i.e.

$$L^{i,I}(\rho) = \ln \int e^{L^i} d\lambda_i.$$

Summing over  $i$  yields the logarithm of the integrated likelihood,

$$L^I(\rho) = \sum_i L^{i,I}(\rho) = \sum_i \ln \int e^{L^i} d\lambda_i. \quad (1)$$

After integrating out the fixed effects, the mode of the integrated likelihood can be used as an estimator<sup>1</sup>. Let the *integrated likelihood estimator*  $\hat{\rho}$  be the mode of  $L^I(\rho)$ ,

$$\hat{\rho} = \arg \max_{\rho} L^I(\rho).$$

A parametrization of the likelihood is *information-orthogonal* if the information matrix is block diagonal. That is

$$EL_{\rho\lambda}(\rho_0, \lambda_0) = 0$$

i.e.

$$\int_{t_{\min}}^{t_{\max}} L_{\rho\lambda}(\rho_0, \lambda_0) e^{L(\rho_0, \lambda_0)} dt = 0,$$

where  $t$  denotes the dependent variable,  $t \in [t_{\min}, t_{\max}]$  and  $\rho_0, \lambda_0$  denote the true value of the parameters. Cox and Reid (1987) and Jeffreys (1961) use this concept and refer to it as ‘orthogonality’. We prefer the term information-orthogonality to distinguish it from the other orthogonality concepts and to stress that it is defined in terms of the properties of the information matrix. See Tibshirani and Wasserman (1994) and Woutersen (2000) for

<sup>1</sup>As  $N \rightarrow \infty$ , using the marginal posteriors is asymptotically equivalent. Considering the mode of the posterior, however, simplifies the algebra.

an overview of orthogonality concepts. Consider the log likelihood  $L(\rho, f(\rho, \lambda))$  where the nuisance parameter  $f$  is written as a function of  $\rho$  and the orthogonal nuisance parameter  $\lambda$ . Differentiating  $L(\rho, f(\rho, \lambda))$  with respect to  $\rho$  and  $\lambda$  yields

$$\begin{aligned}\frac{\partial L(\rho, f(\rho, \lambda))}{\partial \rho} &= L_\rho + L_f \frac{\partial f}{\partial \rho} \\ \frac{\partial^2 L(\rho, f(\rho, \lambda))}{\partial \lambda \partial \rho} &= L_{f\rho} \frac{\partial f}{\partial \lambda} + L_{ff} \frac{\partial f}{\partial \lambda} \frac{\partial f}{\partial \rho} + L_f \frac{\partial^2 f}{\partial \lambda \partial \rho}\end{aligned}$$

where  $L_f$  is a score and therefore  $EL_f = 0$ . Information orthogonality requires that the cross-derivative  $\frac{\partial^2 L(\rho, f(\rho, \lambda))}{\partial \lambda \partial \rho}$  is zero in expectation, i.e.

$$EL_{\rho\lambda} = EL_{f\rho} \frac{\partial f}{\partial \lambda} + EL_{ff} \frac{\partial f}{\partial \lambda} \frac{\partial f}{\partial \rho} = 0.$$

This condition implies the following differential equation

$$EL_{f\rho} + EL_{ff} \frac{\partial f}{\partial \rho} = 0. \quad (2)$$

The information-orthogonal parametrization of the dynamic linear model without regressors is explicit and given by Lancaster (2002).

### 3 Orthogonality in the Dynamic Linear Model with fixed effects

Consider the dynamic linear model with fixed effects,

$$y_{is} = y_{i,s-1}\rho + f_i + \varepsilon_{is} \quad (3)$$

where

$$E(\varepsilon_{is}|y_{i1}, \dots, y_{is}) = 0, \quad E(\varepsilon_{is}^2|y_{i1}, \dots, y_{is}) = \sigma^2, \quad E\varepsilon_{is}\varepsilon_{it}|y_{i1}, \dots, y_{is} = 0$$

for  $s \neq t$ ,  $s = 1, \dots, T$ , and  $i = 1, \dots, N$ . Lancaster (2002) conditions on  $y_{i0}$  and suggests the following parametrization<sup>2</sup>

$$f_i = y_{i0}(1 - \rho) + \lambda_i e^{-b(\rho)} \quad \text{where } b(\rho) = \frac{1}{T} \sum_{s=1}^T \frac{T-s}{s} \rho^s.$$

---

<sup>2</sup>Appendix 5 of Woutersen (2001) gives information-orthogonal parametrizations for linear models with more than one autoregressive term.

Analogue to quasi-maximum likelihood estimators, normality of the error terms is assumed in order to derive the integrated likelihood estimator. The estimator depends only on the first two moments of  $y_{is}$  and is given by Lancaster (2002). Integrating the likelihood contribution of individual  $i$  with respect to  $\lambda$  gives  $e^{L^{i,I}}$ , where

$$e^{L^{i,I}} \propto \frac{1}{\sigma^{T-1}} e^{b(\rho) - \frac{1}{2\sigma^2} \sum_s (y_s - y_{s-1}\rho)^2 + \frac{T}{2} (\overline{y_s - y_{s-1}\rho})^2}.$$

The asymptotic variance can be found by deriving the normalized scores of the integrated likelihood (see appendix 1 for derivation),

$$\begin{aligned} L_\rho^I(\beta, \sigma^2) &= \frac{\sum_i L_\rho^{i,I}}{N} \\ &= b'(\rho) + \frac{1}{\sigma^2} \frac{\sum_i}{N} \left\{ \sum_s (y_s - y_{s-1}\rho) y_{s-1} - \frac{1}{\sigma^2} T (\overline{y_s - y_{s-1}\rho}) \overline{y_{s-1}} \right\} \\ L_{\sigma^2}^I(\beta, \sigma^2) &= \frac{\sum_i L_{\sigma^2}^{i,I}}{N} \\ &= \frac{1}{\sigma^2} \left[ \frac{T-1}{2} - \frac{1}{2\sigma^2} \frac{\sum_i}{N} \left\{ \sum_s (y_s - y_{s-1}\rho)^2 + \frac{T}{2} (\overline{y_s - y_{s-1}\rho})^2 \right\} \right] \end{aligned}$$

where  $b(\rho)' = \frac{1}{T} \sum_{s=1}^T (T-s)\rho^{s-1}$ . In particular, one can show that the difference  $L_\rho^I(\beta_0, \sigma_0^2) - L_\rho(\beta_0, \sigma_0^2, \lambda_0)$  is  $o_p(\frac{1}{\sqrt{NT}})$ . That is, the difference between the score of the integrated likelihood and the score of the regular likelihood is small. Intuitively, this suggests that  $L_\rho^I(\beta, \sigma^2)$  and  $L_\rho(\beta_0, \sigma^2, \lambda_0)$  would yield the similar estimators, say the same up to first order. We show in the next section that this intuition is correct.

## 4 Adaptiveness

Woutersen (2001) shows that the integrated likelihood estimator is adaptive in the sense that it is equivalent to the infeasible maximum likelihood estimator that assumes the nuisance parameters to be known. The conditions for this result are a regularity condition that the integrated likelihood can be approximated by a Laplace formula and the substantial condition that  $T \propto N^\alpha$  where  $\alpha > \frac{1}{3}$ . We now weaken the latter condition for the dynamic linear model. Consider the maximum likelihood estimator for known  $\lambda_0$ . Let  $\{\widehat{\rho}_{ML}, \widehat{\sigma}_{ML}^2\} = \arg \max_\rho L(\rho, \sigma^2, \lambda_0)$ . Assuming normality of the error term, the distribution of  $(\widehat{\rho}_{ML} - \rho)$  is



$\sqrt{NT}(1 - \rho^2)$  with  $T$  increasing and  $N$  constant or increasing<sup>3</sup>. Hahn and Kuersteiner (2002) establish the same bound using a Hajek-type convolution theorem<sup>4</sup>. The following theorem states that the integrated likelihood estimator reaches this theoretical bound.

### Theorem 1

*Let the data be generated by equation (3); let  $|\rho| < 1$  and  $T$  increasing while  $N$  is constant or increasing. Then the integrated likelihood estimator  $\hat{\rho}$  is an adaptive estimator and*

$$\sqrt{NT}(\hat{\rho}_I - \rho_0) \rightarrow_d N(0, 1 - \rho^2),$$

where  $1 - \rho^2$  equals at the Cramér-Rao lower bound.

*Proof:* See appendix.

Efficiency bounds were developed by Stein (1956) and are also discussed in Bickel (1982), Newey (1990) and Bickel et al (1993).

## 5 Simulation Results

We use the same simulation designs as Hahn, Hausman and Kuersteiner (2001), *HHK*, and Hahn and Kuersteiner (2002), *HK*. In particular, the fixed effects  $\alpha_i$  and the innovations  $\varepsilon_{it}$  are assumed to have independent standard normal distributions,  $N(0, 1)$ . Initial observations  $y_{i0}|\alpha_i$  are assumed to be generated by the stationary distribution  $N\left(\frac{\alpha_i}{1-\rho_0}, \frac{\text{Var}(\varepsilon_{it})}{1-\rho_0^2}\right)$ , see section 9 for the tables. In table 1 we consider the same parameter values as *HHK* and calculate the root mean squared error for the integrated likelihood estimator,  $\hat{\rho}_I$ . For convenience, we also reproduce the simulation results of *HHK*;  $\hat{\rho}_{BC2}$  denotes the Bias Corrected estimator of *HHK*,  $\hat{\rho}_{LIML}$  the LIML estimator and  $\hat{\rho}_{GMM}$  the *GMM* estimator of Arellano and Bover (1995). All results are based on 5000 replications and the integrated likelihood estimator has a lower MSE than the other estimators for all parameter values. Table 2 gives the bias of all estimator and the integrated likelihood estimator performs comparable to the other estimators for low values of  $\rho$  and better for higher values. Table 3 gives the simulation design of *HK*. The estimator of *HK* is denoted by  $\hat{\rho}$  and the RMSE of the integrated likelihood

<sup>3</sup>See Lemma 1 of the appendix.

<sup>4</sup>That is,  $N(0, 1 - \rho^2)$  is the minimal asymptotic distribution.

estimator is lower than the RMSE of the other estimators in all cases. In table 4, we compare the integrated likelihood estimator to the ‘long difference’ 2SLS estimator,  $\hat{\rho}_{I2SLS,LD}$ , and ‘long difference’ continuous updating estimator,  $\hat{\rho}_{CUE,LD}$ , of HHK. The RMSE of the integrated likelihood estimator is lower than the RMSE of the other estimators in all cases and the bias is lower in most cases. The lower part of table 4 as well as table 5 show a very good performances of the integrated likelihood estimator in the vicinity of unit root. That is, both in terms of RMSE and bias.

We can summarize our estimation results by saying that the simulation results shows the relevance of our theoretical efficiency result. Our efficiency result allows  $T$  to increase very slowly and, as a result, the integrated likelihood estimator is superior in terms of RMSE and very good in terms of bias.

## 6 Conclusion

This paper considers the dynamic linear model with fixed effects and derives an adaptiveness result for the integrated likelihood estimator. In particular, the integrated likelihood estimator is shown to be adaptive for an asymptotic with  $T$  increasing where  $T$  is the number of observations per individual. Simulations show the relevance of the adaptiveness result. In particular, the Root Mean Squared Error of the integrated likelihood estimator is smaller than the Root Mean Squared Error of competing estimators for any of the parameter values and performs very good in terms of bias. Moreover, the integrated likelihood estimator is consistent for fixed  $T$  and performs very well for  $\rho$  close to unit root.

## 7 Appendices

### Appendix 1.

We assume normality of the error term,  $\varepsilon_{is} \sim N(0, \sigma^2)$ , to derive moment conditions.

Integrating the likelihood with respect to  $\lambda$  gives:

$$\begin{aligned} e^{L^{i,I}}(\rho, \sigma^2) &= \frac{1}{\sigma^T} \int e^L df = \frac{1}{\sigma^T} e^{b(\rho)} \int e^{-\frac{1}{2\sigma^2} \sum_s (y_s - y_{s-1} \rho - f)^2} df \\ &= \frac{1}{\sigma^T} e^{b(\rho) - \frac{1}{2\sigma^2} \sum_s (y_s - y_{s-1} \rho)^2} \int e^{-\frac{T}{2\sigma^2} \{f^2 - 2f \overline{(y_s - y_{s-1} \rho)}\}} df \\ &\propto \frac{1}{\sigma^{T-1}} e^{b(\rho) - \frac{1}{2\sigma^2} \sum_s (y_s - y_{s-1} \rho)^2 + \frac{T}{2} \overline{(y_s - y_{s-1} \rho)}^2}. \end{aligned}$$

This implies the following log of the integrated likelihood contribution for individual  $i$ :

$$L^{i,I} = \frac{1-T}{2} \ln(\sigma^2) + b(\rho) - \frac{1}{2\sigma^2} \sum_{s=1}^T (y_s - y_{s-1} \rho)^2 + \frac{T}{2\sigma^2} \overline{(y_s - y_{s-1} \rho)}^2,$$

where we suppressed the arguments of  $L^{i,I}$ . Differentiating with respect to  $\rho$  and  $\sigma^2$  gives

$$\begin{aligned} L_{\rho}^{i,I} &= b'(\rho) + \frac{1}{\sigma^2} \left\{ \sum_{s=1}^T (y_s - y_{s-1} \rho) y_{s-1} - T \overline{(y_s - y_{s-1} \rho)} \overline{y_{s-1}} \right\} \\ L_{\sigma^2}^{i,I} &= \frac{1}{\sigma^2} \left\{ \frac{1-T}{2} - \frac{1}{2\sigma^2} \sum_{s=1}^T (y_s - y_{s-1} \rho)^2 + \frac{T}{2} \overline{(y_s - y_{s-1} \rho)}^2 \right\}. \end{aligned}$$

We use  $\frac{\sum_i L_{\rho}^{i,I}}{NT}$  and  $\frac{\sum_i L_{\sigma^2}^{i,I}}{NT}$  as moment functions.

### Appendix 2. Lemma 1

To be shown:

$$\sqrt{NT}(\hat{\rho}_{ML} - \rho_0) \rightarrow_d N(0, 1 - \rho^2).$$

*Proof:*

We assume normality of the error term,  $\varepsilon_{is} \sim N(0, \sigma^2)$ , in order to derive the  $ML$  estimator.

The log-likelihood with known orthogonal fixed effects for individual  $i$ ,  $L^i$ , has the following form:

$$L^i = -\frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{s=1}^T (\tilde{y}_{is} - \tilde{y}_{i,s-1} \rho - \lambda_i e^{-b(\rho)})^2 \text{ where } \tilde{y}_{is} = y_{is} - y_{i0}.$$

The asymptotic variance only depends on the first two moments. Without assuming normality, the asymptotic variance of  $ML$  estimator for orthogonal fixed effects has the following form:

$$Asy.var(\hat{\rho}_{ML}) = \left[ \frac{1}{NT} EL_{\rho\rho} \right]^{-1} \left[ \frac{1}{NT} E((L_{\rho})(L_{\rho})') \right] \left[ \frac{1}{NT} EL_{\rho\rho} \right]^{-1}.$$

To derive  $Asy.var(\hat{\rho}_{ML})$ , we start differentiating  $L^i$  with respect to  $\rho$  and  $\sigma^2$ , which gives:

$$\begin{aligned} L_{\rho}^i &= \frac{1}{\sigma^2} \sum_s (\tilde{y}_{is} - \tilde{y}_{i,s-1}\rho - \lambda_i e^{-b(\rho)}) (\tilde{y}_{i,s-1} - \lambda_i b'(\rho) e^{-b(\rho)}) \\ L_{\sigma^2}^i &= -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_s (\tilde{y}_{is} - \tilde{y}_{i,s-1}\rho - \lambda_i e^{-b(\rho)})^2. \end{aligned}$$

Given that  $\lambda_i$  is the orthogonal parametrization, using  $\frac{\partial f_i}{\partial \rho} = -y_0 + \lambda_i b'(\rho) e^{-b(\rho)}$  yields:

$$\begin{aligned} L_{\rho}^i &= \frac{1}{\sigma^2} \sum_s (y_{is} - y_{i,s-1}\rho - f_i) \left( y_{i,s-1} + \frac{\partial f_i}{\partial \rho} \right), \\ L_{\sigma^2}^i &= \frac{-T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{s=1}^T (y_{is} - y_{i,s-1}\rho - f_i)^2. \end{aligned}$$

Equalizing  $L_{\sigma^2} = 0$  gives  $\hat{\sigma}_{ML}^2 = \frac{1}{NT} \sum_i \sum_{s=1}^T (y_{is} - y_{i,s-1}\rho - f_i)^2$ . Considering the fact that  $\sigma^2$  is unknown, we replace it by  $\hat{\sigma}_{ML}^2$  in  $L_{\rho}^i$ . Thus,

$$L_{\rho}^i = \frac{\sum_s (y_{is} - y_{i,s-1}\rho - f_i) \left( y_{i,s-1} + \frac{\partial f_i}{\partial \rho} \right)}{\frac{1}{NT} \sum_i \sum_{s=1}^T (y_{is} - y_{i,s-1}\rho - f_i)^2}.$$

We prove that  $\frac{\partial f_i}{\partial \rho} = -E\bar{y}_{i,s-1}$  when  $T \rightarrow \infty$ :

$$\begin{aligned} \lim_{T \rightarrow \infty} E\bar{y}_{i,s-1} &= \lim_{T \rightarrow \infty} E \frac{\sum_{s=1}^T \left( \rho^{s-1} y_{i0} + (T-s+1) \rho^{s-1} f_i + \varepsilon_{i,s-1} \sum_{j=1}^{T-1} \rho^{j-s} \right)}{T} \\ &= \lim_{T \rightarrow \infty} \frac{\sum_s \rho^{s-1} y_{i0}}{T} + \lim_{T \rightarrow \infty} \frac{\sum_{s=1}^T (T-s+1) \rho^{s-1} f_i}{T} \\ &\quad + \lim_{T \rightarrow \infty} \frac{\sum_{s=1}^T E\varepsilon_{i,s-1} \sum_{j=1}^{T-1} \rho^{j-s}}{T} \\ &= \lim_{T \rightarrow \infty} \frac{1}{1-\rho} \frac{y_{i0}}{T} + \frac{1}{1-\rho} f_i - \lim_{T \rightarrow \infty} \frac{\sum_{s=1}^T (s-1) \rho^{s-1} f_i}{T} \\ &= \frac{1}{1-\rho} f_i + O(T^{-1}), \text{ because } \frac{1}{1-\rho} \frac{y_{i0}}{T} \text{ is } O(T^{-1}) \\ &\quad \text{and } \frac{\sum_{s=1}^T (s-1) \rho^{s-1} f_i}{T} \text{ is } O(T^{-1}). \end{aligned}$$

Thus,  $E\bar{y}_{i,s-1} \rightarrow \frac{f_i}{1-\rho}$  and given that

$$\begin{aligned} \lim_{T \rightarrow \infty} b'(\rho) &= \lim_{T \rightarrow \infty} \left( \frac{1}{T} \sum_{s=1}^T (T-s) \rho^{s-1} \right) = \lim_{T \rightarrow \infty} \sum_{s=1}^T \rho^{s-1} - \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=1}^T s \rho^{s-1} \\ &= \frac{1}{1-\rho} - \lim_{T \rightarrow \infty} \frac{1}{T} \frac{1}{(1-\rho)^2} = \frac{1}{1-\rho} \\ \frac{\partial f_i}{\partial \rho} &\rightarrow -\frac{f_i}{1-\rho} = -E\bar{y}_{i,s-1}, \text{ when } T \rightarrow \infty. \end{aligned}$$

Using  $\varepsilon_{is} = y_{is} - y_{i,s-1}\rho - f_i$  and the fact that  $\frac{\partial f_i}{\partial \rho} = -E\bar{y}_{i,s-1}$ , we have:

$$\begin{aligned} L_\rho^i &= \frac{\sum_s \varepsilon_{is} (y_{i,s-1} - E\bar{y}_{i,-})}{\frac{1}{NT} \sum_i \sum_s \varepsilon_{is}^2} \text{ and differentiating } L_\rho^i \text{ with respect to } \rho \text{ gives} \\ L_{\rho\rho}^i &= \frac{-\sum_s (y_{i,s-1} - E\bar{y}_{i,-})^2}{\frac{1}{NT} \sum_i \sum_s \varepsilon_{is}^2} + \frac{\frac{2}{NT} (\sum_s \varepsilon_{is} (y_{i,s-1} - E\bar{y}_{i,-})) (\sum_i \sum_s \varepsilon_{is} (y_{i,s-1} - E\bar{y}_{i,-}))}{(\frac{1}{NT} \sum_i \sum_s \varepsilon_{is}^2)^2}. \end{aligned}$$

Taking expectations of  $L_{\rho\rho}^i$ , we have:

$$\begin{aligned} EL_{\rho\rho}^i &= \frac{-E \sum_s (y_{i,s-1} - E\bar{y}_{i,-})^2}{\sigma^2} + O((NT)^{-1}) + \frac{2}{NT} \frac{\sum_s E (y_{i,s-1} - E\bar{y}_{i,-})^2}{\sigma^2} + O((NT)^{-1}) \\ &= \frac{-\sum_s E y_{i,s-1}^2 + T (E\bar{y}_{i,s-1})^2}{\sigma^2} + O((NT)^{-1}) + O(N^{-1}), \end{aligned}$$

and thus we derive  $\frac{1}{NT} EL_{\rho\rho}$  as

$$\frac{1}{NT} EL_{\rho\rho} = -\frac{1}{NT\sigma^2} \sum_i \sum_s E y_{i,s-1}^2 + \frac{1}{N\sigma^2} \sum_i (E\bar{y}_{i,s-1})^2 + O(N^{-1}T^{-2}) + O(N^{-1}T^{-1}).$$

We now consider  $(L_\rho^i) (L_\rho^i)'$ :

$$\begin{aligned} (L_\rho^i) (L_\rho^i)' &= \left( \frac{\sum_s \varepsilon_{is} (y_{i,s-1} - E\bar{y}_{i,-})}{\frac{1}{NT} \sum_i \sum_s \varepsilon_{is}^2} \right)^2 \\ &= \frac{\sum_s \varepsilon_{is}^2 (y_{i,s-1} - E\bar{y}_{i,-})^2 + 2 \sum_{j \neq s} (\varepsilon_{is}) (\varepsilon_{ij}) (y_{i,s-1} - E\bar{y}_{i,-}) (y_{i,j-1} - E\bar{y}_{i,-})}{(\frac{1}{NT} \sum_i \sum_s \varepsilon_{is}^2)^2}. \end{aligned}$$

Given that  $E\varepsilon_{is} = 0$ ,  $E\varepsilon_{is}^2 = \sigma^2$  and  $E\varepsilon_{is}\varepsilon_{ij} = 0$  for  $j \neq s$ , we have  $E((L_\rho^i) (L_\rho^i)')$  that is

$$E(L_\rho^i) (L_\rho^i)' = E \frac{\sum_s \varepsilon_{is}^2 (y_{i,s-1} - E\bar{y}_{i,-})^2 + 2 \sum_{j \neq s} (\varepsilon_{is}) (\varepsilon_{ij}) (y_{i,s-1} - E\bar{y}_{i,-}) (y_{i,j-1} - E\bar{y}_{i,-})}{(\frac{1}{NT} \sum_i \sum_s \varepsilon_{is}^2)^2},$$

$$\text{with } E \frac{\sum_s \varepsilon_{is}^2 (y_{i,s-1} - E\bar{y}_{i,-})^2 + 2 \sum_{j \neq s} (\varepsilon_{is}) (\varepsilon_{ij}) (y_{i,s-1} - E\bar{y}_{i,-}) (y_{i,j-1} - E\bar{y}_{i,-})}{(\frac{1}{NT} \sum_i \sum_s \varepsilon_{is}^2)^2} =$$

$$\begin{aligned}
&= \frac{1}{\sigma^2} \sum_s E y_{i,s-1}^2 - \frac{2T}{\sigma^2} (E \bar{y}_{i,s-1})^2 + \frac{T}{\sigma^2} (E \bar{y}_{i,s-1})^2 + O((NT)^{-1}) \\
&= \frac{1}{\sigma^2} \sum_s E y_{i,s-1}^2 - \frac{T}{\sigma^2} (E \bar{y}_{i,s-1})^2 + O((NT)^{-1}).
\end{aligned}$$

Independence across individuals and  $EL_\rho = 0$  gives  $\frac{1}{NT} E((L_\rho)(L_\rho)') = \frac{1}{NT} E \sum_i (L_\rho^i)(L_\rho^i)'$ , so, we have

$$\frac{1}{NT} E((L_\rho)(L_\rho)') = \frac{1}{NT\sigma^2} E \sum_i \sum_s E y_{i,s-1}^2 - \frac{1}{N\sigma^2} \sum_i (E \bar{y}_{i,s-1})^2 + O(N^{-1}T^{-2}),$$

which gives

$$\frac{1}{NT} E((L_\rho)(L_\rho)') = -\frac{1}{NT} EL_{\rho\rho} + O(N^{-1}T^{-1}).$$

And thus the *Asy.var* ( $\hat{\rho}_{ML}$ ) is given by:

$$Asy.var(\hat{\rho}_{ML}) = \left[ -\frac{1}{NT} EL_{\rho\rho} \right]^{-1} + O(N^{-1}T^{-1}).$$

Now using the fact that *Asy.var* ( $\hat{\rho}_{ML}$ ) =  $[-\frac{1}{NT} EL_{\rho\rho}]^{-1} + O(N^{-1}T^{-1})$  we can rewrite it as:

$$\begin{aligned}
Asy.var(\hat{\rho}_{ML}) &= \left[ \frac{1}{NT\sigma^2} \sum_i \sum_{s=1}^T E(y_{i,s-1} - E(\bar{y}_{i-}))^2 \right]^{-1} + O(N^{-1}T^{-1}) \\
&= \left[ \frac{1}{\sigma^2} Var(y_{i,s-1}) \right]^{-1} + O(N^{-1}T^{-1}) = \left[ \frac{\sigma^2}{(1-\rho^2)\sigma^2} \right]^{-1} + O(N^{-1}T^{-1}) \\
Asy.var(\hat{\rho}_{ML}) &= 1 - \rho^2 + O(N^{-1}T^{-1}).
\end{aligned}$$

Thus, the asymptotic variance of the infeasible maximum likelihood estimator equals at the limit the Cramer-Rao lower bound  $(1 - \rho^2)$  (see also Hahn and Kuersteiner (2002))

*Q.E.D.*

### Appendix 3. Theorem 1

To be shown:

$$\sqrt{NT}(\hat{\rho}_I - \rho_0) \rightarrow_d N(0, 1 - \rho^2).$$

*Proof:*

a) Determining the asymptotic variance of the integrated likelihood estimator.

Using  $L_\rho^I$  as a moment, yields the following asymptotic variance (see, for example, Newey and McFadden (1994)).

$$Asy.var(\hat{\rho}_I) = \left[ \frac{1}{NT} EL_{\rho\rho}^I \right]^{-1} \left[ \frac{1}{NT} E((L_\rho^I)(L_\rho^I)') \right] \left[ \frac{1}{NT} EL_{\rho\rho}^I \right]^{-1}.$$

Using the result from the appendix 1 and equalizing  $L_{\sigma^2}^I(\rho, \sigma^2) = 0$  gives

$$\hat{\sigma}_I^2 = \frac{1}{N(T-1)} \sum_i \left( \sum_{s=1}^T (y_{is} - y_{i,s-1}\rho)^2 - T \overline{(y_{is} - y_{i,s-1}\rho)^2} \right).$$

Considering the fact that  $\sigma^2$  is unknown, we replace it by  $\hat{\sigma}_I^2$  in  $L_{\rho}^i$ . This yields

$$L_{\rho}^{i,I} = b'(\rho) + \frac{\sum_s (y_{is} - y_{i,s-1}\rho) y_{i,s-1} - T \overline{(y_{is} - y_{i,s-1}\rho) y_{i,s-1}}}{\frac{1}{N(T-1)} \sum_i \left( \sum_{s=1}^T (y_{is} - y_{i,s-1}\rho)^2 - T \overline{(y_{is} - y_{i,s-1}\rho)^2} \right)}.$$

Using  $\varepsilon_{is} = y_{is} - y_{i,s-1}\rho - f_i$  gives

$$\begin{aligned} L_{\rho}^{i,I} &= b'(\rho) + \frac{\sum_s (\varepsilon_{is} + f_i) y_{i,s-1} - T \overline{(\varepsilon_{is} + f_i) y_{i,-}}}{\frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T \overline{(\varepsilon_{is} + f_i)^2} \right)} \\ &= b'(\rho) + \frac{\sum_s \varepsilon_{is} y_{i,s-1} - T \overline{\varepsilon_i y_{i,-}}}{\frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T \overline{(\varepsilon_{is} + f_i)^2} \right)}, \end{aligned}$$

and differentiating  $L_{\rho}^i$  with respect to  $\rho$  gives

$$\begin{aligned} L_{\rho\rho}^{i,I} &= b''(\rho) + \frac{-\sum_s y_{i,s-1}^2 + T \overline{y_{i,s-1}^2}}{\frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T \overline{(\varepsilon_{is} + f_i)^2} \right)} \\ &\quad + \frac{\frac{2}{N(T-1)} \left( \sum_s \varepsilon_{is} y_{i,s-1} - T \overline{\varepsilon_i y_{i,s-1}} \right) \left( \sum_i \left( \sum_s \varepsilon_{is} y_{i,s-1} - T \overline{\varepsilon_i y_{i,-}} \right) \right)}{\left( \frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T \overline{(\varepsilon_{is} + f_i)^2} \right) \right)^2} \end{aligned}$$

Taking expectations of  $L_{\rho\rho}^{i,I}$  yields:

$$\begin{aligned} EL_{\rho\rho}^{i,I} &= b''(\rho) + E \frac{-\sum_s y_{i,s-1}^2 + T \overline{y_{i,s-1}^2}}{\frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T \overline{(\varepsilon_{is} + f_i)^2} \right)} \\ &\quad + E \frac{\frac{2}{N(T-1)} \left( \sum_s \varepsilon_{is} y_{i,s-1} - T \overline{\varepsilon_i y_{i,s-1}} \right) \left( \sum_i \left( \sum_s \varepsilon_{is} y_{i,s-1} - T \overline{\varepsilon_i y_{i,-}} \right) \right)}{\left( \frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T \overline{(\varepsilon_{is} + f_i)^2} \right) \right)^2} \\ &= b''(\rho) - \frac{1}{\sigma^2} E \sum_s y_{i,s-1}^2 + \frac{T}{\sigma^2} E \overline{y_{i,-}^2} + O((NT)^{-1}) \end{aligned}$$

where  $\frac{1}{NT} EL_{\rho\rho}^I$  is

$$\frac{1}{NT} EL_{\rho\rho}^I = \frac{1}{T} b''(\rho) - \frac{1}{NT\sigma^2} E \left( \sum_i \sum_s y_{i,s-1}^2 \right) + \frac{1}{N\sigma^2} E \left( \sum_i \overline{y_{i,-}^2} \right) + O(N^{-1}T^{-2}),$$

with  $b''(\rho) = \frac{1}{T} \sum_{s=1}^T (T-s)(s-1)\rho^{s-2}$  and

$$\begin{aligned}
\lim_{T \rightarrow \infty} b''(\rho) &= \lim_{T \rightarrow \infty} \left( \frac{1}{T} \sum_{s=1}^T (T-s)(s-1)\rho^{s-2} \right) \\
&= \lim_{T \rightarrow \infty} \left( \frac{1}{T} \sum_{s=1}^T (Ts - T - s^2 + s)\rho^{s-2} \right) \\
&= \lim_{T \rightarrow \infty} \sum_{s=1}^T s\rho^{s-2} - \lim_{T \rightarrow \infty} \sum_{s=1}^T \rho^{s-2} - \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=1}^T s^2 \rho^{s-2} + \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=1}^T s\rho^{s-2} \\
&= \lim_{T \rightarrow \infty} \sum_{s=2}^T (s-1)\rho^{s-2} + \lim_{T \rightarrow \infty} \sum_{s=2}^T \rho^{s-2} - \frac{1}{1-\rho} + O(T^{-1}) \\
&= \frac{1}{(1-\rho)^2} + \frac{1}{1-\rho} - \frac{1}{1-\rho} + O(T^{-1}) = \frac{1}{(1-\rho)^2} + O(T^{-1}),
\end{aligned}$$

$$\frac{1}{NT} EL_{\rho\rho}^I = \frac{1}{T(1-\rho)^2} + O(T^{-2}) - \frac{1}{NT\sigma^2} E \left( \sum_i \sum_s y_{i,s-1}^2 \right) + \frac{1}{N\sigma^2} E \left( \sum_i \bar{y}_{i,-}^2 \right) + O(N^{-1}T^{-2}).$$

Now, considering  $(L_{\rho}^{i,I}) (L_{\rho}^{i,I})'$ , we have:

$$\begin{aligned}
(L_{\rho}^{i,I}) (L_{\rho}^{i,I})' &= \left( b'(\rho) + \frac{\sum_s \varepsilon_{is}(y_{i,s-1} - E\bar{y}_{i,-})}{\frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T(\overline{\varepsilon_{is} + f_i})^2 \right)} \right)^2 \\
&= b'(\rho)^2 + 2b'(\rho) \frac{\sum_s \varepsilon_{is}(y_{i,s-1} - E\bar{y}_{i,-})}{\frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T(\overline{\varepsilon_{is} + f_i})^2 \right)} \\
&\quad + \frac{\sum_s \varepsilon_{is}^2 (y_{i,s-1} - E\bar{y}_{i,-})^2 + 2 \sum_{j \neq s} (\varepsilon_{is})(\varepsilon_{ij})(y_{i,s-1} - E\bar{y}_{i,-})(y_{i,j-1} - E\bar{y}_{i,-})}{\left( \frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T(\overline{\varepsilon_{is} + f_i})^2 \right) \right)^2}.
\end{aligned}$$

Given that  $E\varepsilon_{is} = 0$ ,  $E\varepsilon_{is}^2 = \sigma^2$  and  $E\varepsilon_{is}\varepsilon_{ij} = 0$  for  $j \neq s$ , we have  $E \left( (L_{\rho}^{i,I}) (L_{\rho}^{i,I})' \right)$  that is

$$\begin{aligned}
E(L_{\rho}^{i,I}) (L_{\rho}^{i,I})' &= b'(\rho)^2 + 2b'(\rho) E \frac{\sum_s \varepsilon_{is}(y_{i,s-1} - E\bar{y}_{i,-})}{\frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T(\overline{\varepsilon_{is} + f_i})^2 \right)} \\
&\quad + E \frac{\sum_s \varepsilon_{is}^2 (y_{i,s-1} - E\bar{y}_{i,-})^2 + 2 \sum_{j \neq s} (\varepsilon_{is})(\varepsilon_{ij})(y_{i,s-1} - E\bar{y}_{i,-})(y_{i,j-1} - E\bar{y}_{i,-})}{\left( \frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T(\overline{\varepsilon_{is} + f_i})^2 \right) \right)^2}.
\end{aligned}$$

But

$$E \frac{\sum_s \varepsilon_{is}(y_{i,s-1} - E\bar{y}_{i,-})}{\frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T(\overline{\varepsilon_{is} + f_i})^2 \right)} = 0 + O((NT)^{-1}) \text{ and}$$



$$\begin{aligned}
& E \frac{\sum_s \varepsilon_{is}^2 (y_{i,s-1} - E\bar{y}_{i,-})^2 + 2 \sum_{j \neq s} (\varepsilon_{is})(\varepsilon_{ij})(y_{i,s-1} - E\bar{y}_{i,-})(y_{i,j-1} - E\bar{y}_{i,-})}{\left( \frac{1}{N(T-1)} \sum_i \left( \sum_s (\varepsilon_{is} + f_i)^2 - T(\bar{\varepsilon}_{is} + f_i)^2 \right) \right)^2} \\
&= \frac{1}{\sigma^2} \sum_s E y_{i,s-1}^2 - \frac{2T}{\sigma^2} (E\bar{y}_{i,s-1})^2 + \frac{T}{\sigma^2} (E\bar{y}_{i,s-1})^2 + O\left((NT)^{-1}\right) \\
&= \frac{1}{\sigma^2} \sum_s E y_{i,s-1}^2 - \frac{T}{\sigma^2} (E\bar{y}_{i,s-1})^2 + O\left((NT)^{-1}\right), \text{ so, we have that :}
\end{aligned}$$

$$E(L_\rho^{i,I})(L_\rho^{i,I})' = b'(\rho)^2 + \frac{1}{\sigma^2} \sum_s E y_{i,s-1}^2 - \frac{T}{\sigma^2} (E\bar{y}_{i,s-1})^2 + O\left((NT)^{-1}\right).$$

Independence across individuals and  $EL_\rho^I = 0$  gives  $\frac{1}{NT} E \left( (L_\rho^I) (L_\rho^I)' \right) = \frac{1}{NT} E \sum_i (L_\rho^{i,I}) (L_\rho^{i,I})'$ , so, we have that

$$\frac{1}{NT} E (L_\rho^I) (L_\rho^I)' = \frac{b'(\rho)^2}{T} + \frac{1}{NT\sigma^2} \sum_i \sum_s E y_{i,s-1}^2 - \frac{1}{N\sigma^2} \sum_i (E\bar{y}_{i,s-1})^2 + O(N^{-1}T^{-2}),$$

where  $b'(\rho) = \frac{1}{T} \sum_{s=1}^T (T-s)\rho^{s-1}$ ;  $b'(\rho)^2 = \left( \frac{1}{T} \sum_{s=1}^T (T-s)\rho^{s-1} \right)^2$ , and

$$\begin{aligned}
\lim_{T \rightarrow \infty} b'(\rho)^2 &= \lim_{T \rightarrow \infty} \left( \frac{1}{T} \sum_{s=1}^T (T-s)\rho^{s-1} \right)^2 = \lim_{T \rightarrow \infty} \left( \sum_{s=1}^T \rho^{s-1} - \frac{1}{T} \sum_{s=1}^T s\rho^{s-1} \right)^2 \\
&= \lim_{T \rightarrow \infty} \left( \sum_{s=1}^T \rho^{s-1} \right)^2 - \lim_{T \rightarrow \infty} \frac{2}{T} \sum_{s=1}^T \rho^{s-1} \sum_{s=1}^T s\rho^{s-1} + \lim_{T \rightarrow \infty} \left( \frac{1}{T} \sum_{s=1}^T s\rho^{s-1} \right)^2 \\
&= \frac{1}{(1-\rho)^2} + O(T^{-1}) + O(T^{-2}).
\end{aligned}$$

$$\frac{1}{NT} E (L_\rho^I) (L_\rho^I)' = \frac{1}{T(1-\rho)^2} + O(T^{-2}) + \frac{1}{NT\sigma^2} \sum_i \sum_s E y_{i,s-1}^2 - \frac{1}{N\sigma^2} \sum_i (E\bar{y}_{i,s-1})^2 + O(N^{-1}T^{-2})$$

Thus, we get  $\frac{1}{NT} E (L_\rho^I) (L_\rho^I)' = -\frac{1}{NT} EL_{\rho\rho}^I$  that is  $O(T^{-1})$ . This means that, when  $T \rightarrow \infty$ , we can write:

$$\frac{1}{NT} EL_{\rho\rho}^I = -\frac{1}{NT} E \left( (L_\rho^I) (L_\rho^I)' \right) + O(T^{-1}),$$

therefore, the asymptotic variance of the integrated likelihood estimator is given by

$$Asy.var(\hat{\rho}_I) = \left[ -\frac{1}{NT} EL_{\rho\rho}^I \right]^{-1} + O(T^{-1}).$$

$$\begin{aligned}
Asy.var(\hat{\rho}_I) &= \left[ \frac{1}{NT\sigma^2} \sum_i^N \sum_{s=1}^T E(y_{i,s-1} - E(\bar{y}_{i-}))^2 - \frac{1}{T(1-\rho)^2} \right]^{-1} + O(T^{-1}) \\
&= \left[ \frac{1}{\sigma^2} Var(y_{i,s-1}) - \frac{1}{T(1-\rho)^2} \right]^{-1} + O(T^{-1}) \\
&= \left[ \frac{\sigma^2}{(1-\rho^2)\sigma^2} + O(T^{-1}) \right]^{-1} + O(T^{-1}) \\
Asy.var(\hat{\rho}_I) &= 1 - \rho^2 + O(T^{-1}).
\end{aligned}$$

Thus, when  $T \rightarrow \infty$ , the asymptotic variance equals at the limit the Cramer-Rao lower bound,  $(1 - \rho^2)$ .

*Q.E.D.*

b) Integrated likelihood estimator is normally distributed

To prove normality of the integrated likelihood estimator we redefine  $L_\rho^{i,I}$  as:

$$L_\rho^{i,I} = b'(\rho) + \frac{1}{\sigma^2} \sum_t (\varepsilon_{it} - \bar{\varepsilon}_i) y_{i,t-1},$$

where  $\bar{\varepsilon}_i = \frac{\sum_{t=1}^T \varepsilon_{it}}{T}$ . If we define  $u_{it} = (\varepsilon_{it} - \bar{\varepsilon}_i) y_{i,t-1}$ , then

$$L_\rho^{i,I} = b'(\rho) + \frac{1}{\sigma^2} \sum_t u_{it}.$$

If we define  $v_{it} = \frac{b'(\rho)}{T} + \frac{1}{\sigma^2} u_{it}$ , then we check the conditions of the Lindberg-Levy Central Limit Theorem (see Greene (2000)):

$$Ev_{it} = 0$$

$$Ev_{it}^2 < \infty$$

$$Ev_{it} v_{is} < \infty, \text{ for } s \neq t.$$

To check the first condition we consider computing:

$$Ev_{it} = \frac{b'(\rho)}{T} + \frac{1}{\sigma^2} Eu_{it}.$$

Given that  $Eu_{it} = E(\varepsilon_{it} - \bar{\varepsilon}_i) y_{i,t-1}$ , where  $y_{i,t-1} = \rho^{t-1} y_0 + \frac{1-\rho^{t-1}}{1-\rho} f_i + \rho^{t-2} \varepsilon_1 + \dots + \rho \varepsilon_{t-2} + \varepsilon_{t-1}$ ,

and  $E\varepsilon_t = 0$ , we have that

$$\begin{aligned}
Eu_t &= E(\varepsilon_t - \bar{\varepsilon})y_{t-1} \\
&= E(\varepsilon_t - \bar{\varepsilon}) \left( \rho^{t-1}y_0 + \frac{1-\rho^{t-1}}{1-\rho}f_i + \rho^{t-2}\varepsilon_1 + \dots + \rho\varepsilon_{t-2} + \varepsilon_{t-1} \right) \\
&= \rho^{t-1}y_0E(\varepsilon_t - \bar{\varepsilon}) + \frac{1-\rho^{t-1}}{1-\rho}f_iE(\varepsilon_t - \bar{\varepsilon}) \\
&\quad + E(\varepsilon_t - \bar{\varepsilon}) (\rho^{t-2}\varepsilon_1 + \dots + \rho\varepsilon_{t-2} + \varepsilon_{t-1}) \\
&= E(\varepsilon_t - \frac{\varepsilon_1 + \dots + \varepsilon_{T-1} + \varepsilon_T}{T}) (\rho^{t-2}\varepsilon_1 + \dots + \rho\varepsilon_{t-2} + \varepsilon_{t-1}) \\
&= -\frac{\sigma^2}{T} (\rho^{t-2} + \dots + \rho + 1) = -\frac{\sigma^2}{T} \frac{1-\rho^{t-1}}{1-\rho} = -\frac{\sigma^2}{T} b'(\rho)
\end{aligned}$$

and that  $Ev_t = -\frac{\sigma^2 b'(\rho)}{T}$ , we have

$$Ev_t = \frac{b'(\rho)}{T} - \frac{\sigma^2 b'(\rho)}{\sigma^2 T} = 0.$$

If we compute  $Ev_t^2$ , we have

$$\begin{aligned}
Ev_t^2 &= E \left( \frac{b'(\rho)}{T} + \frac{1}{\sigma^2} u_t \right)^2 \\
&= E \left( \left( \frac{b'(\rho)}{T} \right)^2 + 2 \frac{b'(\rho)}{T} \frac{1}{\sigma^2} u_t + \frac{1}{\sigma^4} u_t^2 \right) \\
&= \left( \frac{b'(\rho)}{T} \right)^2 + 2 \frac{b'(\rho)}{T} \frac{1}{\sigma^2} Eu_t + \frac{1}{\sigma^4} Eu_t^2.
\end{aligned}$$

Given that  $\bar{\varepsilon}$  is  $O_p(T^{-\frac{1}{2}})$ , then for  $Eu_t^2$  we have

$$\begin{aligned}
Eu_t^2 &= E[(\varepsilon_t - \bar{\varepsilon})y_{t-1}]^2 = E[(\varepsilon_t y_{t-1} - \bar{\varepsilon} y_{t-1})^2] = E(\varepsilon_t y_{t-1})^2 - 2E\bar{\varepsilon}\varepsilon_t y_{t-1}^2 + E\bar{\varepsilon}^2 y_{t-1}^2 \\
&= E \left[ \varepsilon_t^2 \left( \rho^{t-1}y_0 + \frac{1-\rho^{t-1}}{1-\rho}f_i + \rho^{t-2}\varepsilon_1 + \dots + \rho\varepsilon_{t-2} + \varepsilon_{t-1} \right)^2 \right] + O(T^{-\frac{1}{2}}) + O(T^{-1}) \\
&= E\varepsilon_t^2 [(\rho^{t-1}y_0)^2 + \left( \frac{1-\rho^{t-1}}{1-\rho}f_i \right)^2 + (\rho^{t-2}\varepsilon_1 + \dots + \rho\varepsilon_{t-2} + \varepsilon_{t-1})^2 \\
&\quad + 2(\rho^{t-1}y_0) \left( \frac{1-\rho^{t-1}}{1-\rho}f_i \right) + 2(\rho^{t-1}y_0) (\rho^{t-2}\varepsilon_1 + \dots + \rho\varepsilon_{t-2} + \varepsilon_{t-1}) \\
&\quad + 2 \left( \frac{1-\rho^{t-1}}{1-\rho}f_i \right) (\rho^{t-2}\varepsilon_1 + \dots + \rho\varepsilon_{t-2} + \varepsilon_{t-1})] + O(T^{-\frac{1}{2}}) + O(T^{-1}) \\
&= \sigma^2 \left[ (\rho^{t-1}y_0)^2 + \left( \frac{1-\rho^{t-1}}{1-\rho}f_i \right)^2 + 2(\rho^{t-1}y_0) \left( \frac{1-\rho^{t-1}}{1-\rho}f_i \right) \right] + O(T^{-\frac{1}{2}}) + O(T^{-1}) \\
&= \sigma^2 \left( \rho^{t-1}y_0 + \frac{1-\rho^{t-1}}{1-\rho}f_i \right)^2 + O(T^{-\frac{1}{2}}) + O(T^{-1}).
\end{aligned}$$

Given that  $Eu_t < \infty$  and  $Eu_t^2 < \infty$  for  $|\rho| < 1$  and  $T \rightarrow \infty$ , we have that  $Eu_t^2 < \infty$ .

To check  $Eu_t v_s v_l < \infty$ , for  $s \neq l$  we write

$$\begin{aligned}
Eu_t v_s v_l &= E \left( \frac{b'(\rho)}{T} + \frac{1}{\sigma^2} u_t \right) \left( \frac{b'(\rho)}{T} + \frac{1}{\sigma^2} u_s \right) \left( \frac{b'(\rho)}{T} + \frac{1}{\sigma^2} u_l \right) \\
&= E \left( \frac{b'(\rho)}{T} + \frac{1}{\sigma^2} u_t \right) \left( \left( \frac{b'(\rho)}{T} \right)^2 + \frac{b'(\rho)}{\sigma^4 T} (u_s + u_l) + \frac{1}{\sigma^4} u_s u_l \right) \\
&= \left( \frac{b'(\rho)}{T} \right)^3 + \left( \frac{b'(\rho)}{T} \right)^2 \frac{1}{\sigma^2} E(u_t + u_s + u_l) \\
&\quad + \frac{b'(\rho)}{T \sigma^4} E(u_s u_l + u_t u_s + u_t u_l) + \frac{1}{\sigma^6} E u_t u_s u_l.
\end{aligned}$$

We have that  $E(u_t + u_s + u_l) = -\frac{3\sigma^2}{T} b'(\rho)$ , which is  $O(T^{-1})$ , also we have

$$\begin{aligned}
E(u_s u_l + u_t u_s + u_t u_l) &= E(\varepsilon_s - \bar{\varepsilon})(\varepsilon_l - \bar{\varepsilon})y_{s-1}y_{l-1} \\
&\quad + E(\varepsilon_t - \bar{\varepsilon})(\varepsilon_s - \bar{\varepsilon})y_{t-1}y_{s-1} + E(\varepsilon_t - \bar{\varepsilon})(\varepsilon_l - \bar{\varepsilon})y_{t-1}y_{l-1} \\
&= O(T^{-\frac{1}{2}}) + O(T^{-1}) \text{ when } t \neq s \neq l \text{ and is} \\
&= \sigma^2 \left( \rho^{t-1} y_0 + \frac{1 - \rho^{t-1}}{1 - \rho} f_i \right)^2 + O(T^{-\frac{1}{2}}) + O(T^{-1}) \text{ when } t = s \text{ or } l.
\end{aligned}$$

For  $Eu_t u_s u_l$  we have

$$\begin{aligned}
Eu_t u_s u_l &= E(\varepsilon_t - \bar{\varepsilon})(\varepsilon_s - \bar{\varepsilon})(\varepsilon_l - \bar{\varepsilon})y_{t-1}y_{s-1}y_{l-1} \\
&= E\varepsilon_t \varepsilon_s \varepsilon_l y_{t-1} y_{s-1} y_{l-1} - E\bar{\varepsilon} \varepsilon_t \varepsilon_s y_{t-1} y_{s-1} y_{l-1} - E\bar{\varepsilon} \varepsilon_s \varepsilon_l y_{t-1} y_{s-1} y_{l-1} \\
&\quad - E\bar{\varepsilon} \varepsilon_t \varepsilon_l y_{t-1} y_{s-1} y_{l-1} + E\bar{\varepsilon}^2 (\varepsilon_t + \varepsilon_s + \varepsilon_l) y_{t-1} y_{s-1} y_{l-1} - E\bar{\varepsilon}^3 y_{t-1} y_{s-1} y_{l-1} \\
&= O(T^{-\frac{1}{2}}) + O(T^{-1}) + O(T^{-\frac{3}{2}}).
\end{aligned}$$

Thus,

$$\begin{aligned}
Eu_t v_s v_l &= -2 \left( \frac{b'(\rho)}{T} \right)^3 + \frac{b'(\rho)}{T \sigma^4} \left( O(T^{-\frac{1}{2}}) + O(T^{-1}) \right) \\
&\quad + O(T^{-\frac{1}{2}}) + O(T^{-1}) + O(T^{-\frac{3}{2}}) \\
&= O(T^{-3}) + O(T^{-\frac{1}{2}}) + O(T^{-1}) + O(T^{-\frac{3}{2}}) + O(T^{-2}),
\end{aligned}$$

when  $t \neq s \neq l$  and is

$$\begin{aligned} Ev_t v_s v_l &= -2 \left( \frac{b'(\rho)}{T} \right)^3 + \frac{b'(\rho)}{T\sigma^4} \left( \sigma^2 \left( \rho^{t-1} y_0 + \frac{1-\rho^{t-1}}{1-\rho} f_i \right)^2 + O(T^{-\frac{1}{2}}) + O(T^{-1}) \right) \\ &\quad + O(T^{-\frac{1}{2}}) + O(T^{-1}) + O(T^{-\frac{3}{2}}) \\ &= O(T^{-3}) + O(T^{-1}) + O(T^{-\frac{1}{2}}) + O(T^{-1}) + O(T^{-\frac{3}{2}}) + O(T^{-2}), \end{aligned}$$

when  $t = s$  or  $l$ . Thus, we have that  $Ev_t v_s v_l < \infty$  for  $|\rho| < 1$  and  $T \rightarrow \infty$ .

The conditions of Central Limit Theorem are satisfied for each individual  $i$ . In order to show that the Integrated Likelihood Estimator is asymptotically Normal Distributed we need to account for all individuals  $i$ . Let

$$\frac{L_\rho^{i,l}}{\sqrt{T}} = \varepsilon_i + u_i,$$

where  $\varepsilon_i \sim N\left(0, \frac{EL_{\rho\rho}^{i,l}}{T}\right)$ ,  $Eu_i = 0$  and  $Eu_i^2 = o(1)$ , because we proved that

$$E\left(\frac{L_\rho^{i,l}}{\sqrt{T}}\right) \left(\frac{L_\rho^{i,l}}{\sqrt{T}}\right)' = -E\left(\frac{L_{\rho\rho}^{i,l}}{T}\right) + o(1).$$

Thus we have that:

$$\frac{\sum_i \left(\frac{L_\rho^{i,l}}{\sqrt{T}}\right)}{\sqrt{N}} = \frac{\sum_i \varepsilon_i}{\sqrt{N}} + \frac{\sum_i u_i}{\sqrt{N}}.$$

If we define  $\varepsilon = \frac{\sum_i \varepsilon_i}{\sqrt{N}}$  and  $u = \frac{\sum_i u_i}{\sqrt{N}}$ , then  $\varepsilon \sim N\left(0, \frac{EL_{\rho\rho}^l}{NT}\right)$  and  $E(u) = \frac{\sum_i u_i}{\sqrt{N}} = 0$  and  $E(u^2) = \frac{\sum_i u_i^2}{N} = o(1)$  since we assumed independence across individuals.

We proved that,

$$\frac{L_\rho^l}{\sqrt{NT}} \sim N\left(0, \frac{EL_{\rho\rho}^l}{NT}\right)$$

and the result follows.

*Q.E.D.*

## 8 References

- Alvarez, J. and M. Arellano (1998): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," unpublished manuscript.
- Ahn, S. C. and Schmidt, P. (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5-27.
- Anderson, T. W. and C. Hsiao (1982): "Formulation and estimation of 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. and Bover, O. (1995): "Another Look at the Instrumental Variable Estimation of Error Component Models." *Journal of Econometrics* 68: 29-51.
- Arellano, M., and B. E. Honoré, (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics*, Vol. 5, ed. by J. Heckman and E. Leamer, Amsterdam: North-Holland.
- Baltagi, B. H. (1995): "*Econometric Analysis of Panel Data*," John Wiley and Sons, New York.
- Berger, J. O., B. Liseo, and R. L. Wolpert (1999): "Integrated Likelihood Methods for Eliminating Nuisance Parameters," *Statistical Science*, 14, 1-28.
- Bickel, P. J. (1982), "On Adaptive estimation", *Annals of Statistics*, vol.10, 647 - 671.
- Bickel P. J., C. A. J. Klaassen, Y. Ritov and J. A. Wellner (1993), *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins series in mathematical science.
- Blundell, R. and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*, 87, 115-143.
- Chamberlain, G. (1984): "Panel Data," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland.
- (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 3, 567-596.

- Cox, D. R., and N. Reid (1987): "Parameter Orthogonality and Approximate Conditional Inference (with Discussion)," *Journal of the Royal Statistical Society, Series B*, 49, 1-39.
- (1993): "A Note on the Calculation of Adjusted Profile Likelihood," *Journal of the Royal Statistical Society, Series B*, 45, 467-471.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Rubin (1995): *Bayesian Data Analysis*, Chapman and Hall.
- Greene, W.H. (2000): "Econometric Analysis," Fourth Edition, Prentice Hall, 116-117.
- Hahn, J., J. Hausman, and G. Kuersteiner (2001): "Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects", MIT working paper.
- Hahn, J. and G. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both  $n$  and  $T$  are Large", *Econometrica*, 70, 1639-1658.
- Heckman, J. J. (1981a): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. F. Manski and D. McFadden, pp. 114-178, MIT Press, Cambridge.
- Heckman, J. J. (1981b): "The Incidental Parameter Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. F. Manski and D. McFadden, pp. 179-195, MIT Press, Cambridge.
- Holtz-Eakin, D., Newey, W. and Rosen, H. S. (1988): "Estimating Vector Autoregressions With Panel Data," *Econometrica*, Vol. 56, No. 6, S. 1371-1395.
- Hsiao, C. (1986): *Analysis of Panel Data*. Cambridge University Press, New York
- Hsiao, C., M. H. Pesaran and A. K. Tahmiscioglu (2002): "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods", *Journal of Econometrics*, 1, 67-106.
- Jeffreys, H. (1961): *Theory of Probability*, 3rd ed. Oxford: Clarendon Press.
- Lancaster, T. (2000): "The Incidental Parameters since 1948," *Journal of Econometrics*,

- 95, 391-413.
- Lancaster, T. (2002): "Orthogonal Parameters and Panel Data, *Review of Economic Studies*.
- Mankiw, N. G., Romer D., and D. Weil: "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, CVII, 1992, 407-437.
- Neyman, J., and E. L. Scott (1948): "Consistent Estimation from Partially Consistent Observations," *Econometrica*, 16, 1-32.
- Newey, W. K and R. J. Smith (2001): "Higher Order Properties of *GMM* and Generalized Empirical Likelihood Estimators," Working Paper MIT.
- Newey, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99-135.
- Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4, ed. by R. F. Engle and D. MacFadden. Amsterdam: North-Holland.
- Nickell, S. J. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 48, 879-897.
- Solow, R. M. (1956): "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, LXX, 1956, 65-94.
- Stein, C (1956): "Efficient Nonparametric Testing and Estimation," *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, Vol.1.
- Tibshirani, R., and L. Wasserman (1994): "Some Aspects of Reparametrization of Statistical Models," *The Canadian Journal of Statistics*, Vol. 22, 163-173.
- Wansbeek, T. J. and P. A. Bekker (1996): "On IV, GMM and ML in a dynamic panel data model," *Economics Letters* 51, 145-152.
- Woutersen, T. M. (2000): "Consistent Estimation and Orthogonality," UWO working paper.
- Woutersen, T. M. (2001): "Robustness against Incidental Parameters and Mixing Distributions," UWO working paper.



## 9 Tables

Table 1 The Root Mean Squared Error (RMSE) of the integrated likelihood estimator compared to estimators considered in Hahn, Hausman and Kuersteiner (2001) ( $\sigma^2$  is unknown; 5000 data sets)

$T$	$n$	$\rho$	RMSE $\hat{\rho}_{GMM}$	RMSE $\hat{\rho}_{BC2}$	RMSE $\hat{\rho}_{LIML}$	RMSE $\hat{\rho}_I$
5	100	0.10	0.08	0.08	0.082	0.056
10	100	0.10	0.05	0.05	0.045	0.035
5	500	0.10	0.04	0.04	0.036	0.025
10	500	0.10	0.02	0.02	0.020	0.016
5	100	0.30	0.10	0.10	0.099	0.061
10	100	0.30	0.05	0.05	0.050	0.036
5	500	0.30	0.04	0.04	0.044	0.027
10	500	0.30	0.02	0.02	0.023	0.016
5	100	0.50	0.13	0.13	0.130	0.067
10	100	0.50	0.06	0.06	0.058	0.037
5	500	0.50	0.06	0.06	0.057	0.031
10	500	0.50	0.03	0.03	0.026	0.016
5	100	0.80	0.32	0.34	0.327	0.089
10	100	0.80	0.14	0.11	0.109	0.045
5	500	0.80	0.13	0.13	0.127	0.036
10	500	0.80	0.05	0.04	0.044	0.018
5	100	0.90	0.55	0.78	0.604	0.099
10	100	0.90	0.25	0.23	0.229	0.057
5	500	0.90	0.28	0.30	0.277	0.038
10	500	0.90	0.10	0.08	0.080	0.023

For comparison, the simulation results of HHK are shown;  $\hat{\rho}_{GMM}$  denotes the *GMM* estimator of Arellano and Bover (1995),  $\hat{\rho}_{BC2}$  the Bias Corrected estimator of *HHK*,  $\hat{\rho}_{LIML}$  the LIML estimator,  $\hat{\rho}_{BC2}$  denotes the Bias Corrected estimator of *HHK*, and  $\hat{\rho}_I$  the integrated likelihood estimator.

Table 2 Bias of the integrated likelihood estimator  
 compared to estimators considered in Hahn, Hausman and Kuersteiner (2001)  
 ( $\sigma^2$  is unknown; 5000 data sets)

$T$	$n$	$\rho$	%bias $\hat{\rho}_{GMM}$	%bias $\hat{\rho}_{BC2}$	%bias $\hat{\rho}_{LIML}$	%bias $\hat{\rho}_I$
5	100	0.10	-14.96	0.25	-3	2.66
10	100	0.10	-14.06	-0.77	-1	0.45
5	500	0.10	-3.68	-0.77	-1	-0.20
10	500	0.10	-3.15	-0.16	-1	-0.74
5	100	0.30	-8.86	-0.47	-3	1.04
10	100	0.30	-7.06	-0.66	-1	-0.25
5	500	0.30	-2.03	-0.16	-1	-0.37
10	500	0.30	-1.58	-0.10	0	-0.14
5	100	0.50	-10.05	-1.14	-3	0.38
10	100	0.50	-6.76	-0.93	-1	-0.15
5	500	0.50	-2.25	-0.15	-1	-0.29
10	500	0.50	-1.53	-0.11	0	-0.19
5	100	0.80	-27.65	-11.33	-15	0.80
10	100	0.80	-13.45	-4.55	-5	0.25
5	500	0.80	-6.98	-0.72	-3	-0.12
10	500	0.80	-3.48	-0.37	-1	-0.06
5	100	0.90	-50.22	-42.10	-41	1.36
10	100	0.90	-24.27	-15.82	-15	0.83
5	500	0.90	-20.50	-6.23	-10	-0.07
10	500	0.90	-8.74	-2.02	-2	-0.07

The fixed effects  $\alpha_i$  and the innovations  $\varepsilon_{it}$  are assumed to have independent standard normal distributions. Initial observations  $y_{i0}$  are assumed to be generated by the stationary distribution  $N\left(\frac{\alpha_i}{1-\rho_0}, \frac{1}{1-\rho_0^2}\right)$ . As in table 1,  $\hat{\rho}_{GMM}$  denotes the *GMM* estimator of Arellano and Bover (1995),  $\hat{\rho}_{BC2}$  the Bias Corrected estimator of *HHK*,  $\hat{\rho}_{LIML}$  the *LIML* estimator,  $\hat{\rho}_{BC2}$  the Bias Corrected estimator of *HHK*, and  $\hat{\rho}_I$  the integrated likelihood estimator.

Table 3 The performance of the integrated likelihood estimator compared to the *GMM* and *BCML* estimator considered by Hahn and Kuersteiner (2002) ( $\sigma^2$  is unknown; 5000 data sets)

$T$	$n$	$\rho$	bias $\hat{\rho}_{GMM}$	bias $\hat{\rho}$	bias $\hat{\rho}_I$	RMSE $\hat{\rho}_{GMM}$	RMSE $\hat{\rho}$	RMSE $\hat{\rho}_I$
5	100	0.0	-0.011	-0.039	-0.0004	0.074	0.065	0.054
5	100	0.3	-0.027	-0.069	0.003	0.099	0.089	0.061
5	100	0.6	-0.074	-0.115	0.002	0.160	0.129	0.070
5	100	0.9	-0.452	-0.178	0.012	0.552	0.187	0.099
5	200	0.0	-0.006	-0.041	-0.001	0.053	0.055	0.038
5	200	0.3	-0.014	-0.071	0.001	0.070	0.081	0.042
5	200	0.6	-0.038	-0.116	0.003	0.111	0.124	0.048
5	200	0.9	-0.337	-0.178	0.006	0.443	0.183	0.067
10	100	0.0	-0.011	-0.010	-0.001	0.044	0.036	0.035
10	100	0.3	-0.021	-0.019	-0.0007	0.053	0.040	0.036
10	100	0.6	-0.045	-0.038	0.001	0.075	0.051	0.037
10	100	0.9	-0.218	-0.079	0.007	0.248	0.085	0.058
10	200	0.0	-0.006	-0.011	-0.0007	0.031	0.027	0.025
10	200	0.3	-0.011	-0.019	-0.001	0.038	0.032	0.026
10	200	0.6	-0.025	-0.037	-0.0001	0.051	0.045	0.027
10	200	0.9	-0.152	-0.079	0.004	0.181	0.082	0.041
20	100	0.0	-0.011	-0.003	-0.000	0.029	0.024	0.023
20	100	0.3	-0.017	-0.005	0.000	0.033	0.024	0.023
20	100	0.6	-0.029	-0.011	0.000	0.042	0.024	0.022
20	100	0.9	-0.100	-0.032	0.0005	0.109	0.037	0.026
20	200	0.0	-0.006	-0.003	-0.0002	0.020	0.017	0.017
20	200	0.3	-0.009	-0.005	-0.0002	0.022	0.017	0.017
20	200	0.6	-0.016	-0.010	-0.0004	0.027	0.018	0.015
20	200	0.9	-0.065	-0.031	0.0008	0.074	0.034	0.018

$\hat{\rho}_{GMM}$  denotes the *GMM* estimator of Arellano and Bover (1995),  $\hat{\rho}$  the Bias Corrected estimator of *HK*, and  $\hat{\rho}_I$  the integrated likelihood estimator.

Table 4. The Performance of Integrated likelihood estimator for high values of  $\rho$  and  $T = 5$

$N = 100$		$\hat{\rho}_{LIML,1}$	$\hat{\rho}_{I2SLS,LD}$	$\hat{\rho}_{CUE,LD}$	$\hat{\rho}_I$
$\rho = 0.75$	Actual mean % Bias	1.297	5.533	11.553	1.158
$\rho = 0.75$	Actual median % Bias	-3.087	1.381	7.470	0.475
$\rho = 0.75$	RMSE	0.181	0.176	0.213	0.084
$\rho = 0.80$	Actual mean % Bias	-0.112	4.304	10.413	0.802
$\rho = 0.80$	Actual median % Bias	-5.725	1.457	8.651	-0.112
$\rho = 0.80$	RMSE	0.213	0.173	0.205	0.089
$\rho = 0.85$	Actual mean % Bias	-3.899	1.966	7.983	0.995
$\rho = 0.85$	Actual median % Bias	-10.117	0.065	7.558	-0.022
$\rho = 0.85$	RMSE	0.233	0.160	0.194	0.093
$\rho = 0.90$	Actual mean % Bias	-9.757	-0.771	6.138	1.363
$\rho = 0.90$	Actual median % Bias	-15.389	-2.346	6.114	0.288
$\rho = 0.90$	RMSE	0.246	0.153	0.180	0.099
$\rho = 0.95$	Actual mean % Bias	-15.203	-3.367	3.124	1.455
$\rho = 0.95$	Actual median % Bias	-19.637	-4.776	3.136	-0.045
$\rho = 0.95$	RMSE	0.252	0.149	0.165	0.111

$\hat{\rho}_{LIML,1}$  denotes the LIML estimator,  $\hat{\rho}_{I2SLS,LD}$  the 'long difference' 2SLS estimator of HHK,  $\hat{\rho}_{CUE,LD}$  the 'long difference' continuous updating estimator of HHK and  $\hat{\rho}_I$  the integrated likelihood estimator.

Table 5 The performance of the integrated likelihood estimator  
for  $\rho$  close to one and  $T = \{5, 10\}$ .  
( $\sigma^2$  is unknown; 5000 data sets)

$T$	$n$	$\rho$	RMSE $_{\hat{\rho}_I}$	%bias $\hat{\rho}_I$
5	100	0.95	0.111	1.45
10	100	0.95	0.059	0.78
5	500	0.95	0.042	-0.05
10	500	0.95	0.026	0.02
5	100	0.99	0.118	1.69
10	100	0.99	0.058	0.44
5	500	0.99	0.044	0.15
10	500	0.99	0.028	0.22