

# Department of Economics

## GMM Estimation of Dynamic Panel Data Models with Persistent Data

Hugo Kruiniger

Working Paper No. 428

December 2000

ISSN 1473-0278



# GMM estimation of dynamic panel data models with persistent data.

Hugo Kruiniger\*  
Queen Mary, University of London

This version October 2000.

JEL classification: C11, C14, C23.

Keywords: dynamic panel data models, fixed effects, Generalized Method of Moments, weak moment conditions, local-to-zero asymptotics, local-to-unity asymptotics, redundancy, unit root tests, superefficiency.

---

\*Address: H.KRUINIGER@QMW.AC.UK; Dept. of Economics, Mile End Road, London E1 4NS. This paper has been part of a larger paper entitled “ML and GMM estimation of dynamic panel data models with fixed effects” which was based on research that I started in 1997. Versions of the larger paper have been presented at the Universities of Exeter and Cambridge, the econometrics seminar of LSE-STICERD, Queen Mary and Westfield College, the University of Southampton, ESEM99 in Santiago de Compostella, and at the 2000 ESRC Econometrics Study Group Conference in Bristol. I should like to thank Richard Blundell, Grant Hillier, Jerzy Szroeter, and Elias Tzavalis for helpful comments and stimulating discussions. All errors are solely mine. This research was partly supported by the European Commission through Grant ERBFMBICT961594.

## Abstract

This paper considers GMM based estimation and testing procedures for two versions of the AR(1) model with Fixed Effects, henceforth abbreviated as ARFE(1): the conditional ARFE(1) model, and the inclusive ARFE(1) model, which contains the stationary ARFE(1) models and the ARFE(1) model with a unit root.

First, the paper presents a two-step Optimal Linear GMM (OLGMM) estimator for the inclusive model which is asymptotically equivalent to the optimal nonlinear GMM estimator of Ahn and Schmidt (1997). Then the paper examines the properties of the GMM estimators for both versions of the model when the data are persistent. Among other things, we find that the OLGMM estimator is superefficient in the unit root case. Furthermore, under stationarity the covariances of the instruments of the Arellano-Bond estimator and the first differences of the dependent variable are *not* weak. We also derive new approximations to the finite sample distributions of the Arellano-Bond estimator (for both versions of the model), the Arellano-Bover estimator, and the System estimator. We employ local-to-zero asymptotics (cf Staiger and Stock (1997)) for the Arellano-Bond estimator for the conditional model, because its instruments *are* weak in this context, and we employ local-to-unity asymptotics, which is developed in this paper, for the estimators for the stationary model. The new approximations agree well with the Monte Carlo evidence in terms of bias and variance. Finally, various GMM based unit root tests against stationary and conditional alternatives are proposed.

# 1 Introduction

This paper studies GMM based estimation and testing procedures for two versions of the ARFE(1) model with persistent data: an ARFE(1) model that conditions on the initial observations irrespective of the value of the autoregressive parameter,  $r$ , and an ARFE(1) model that only conditions on the initial observations when the series contain a unit root but otherwise assumes stationarity. As the model of the latter type is the closure of the class of stationary ARFE(1) models, we will refer to it as the inclusive ARFE(1) model. The former type of model will be referred to as the conditional ARFE(1) model.

In keeping with the traditional panel data literature, the asymptotic results in the paper are derived by letting the cross-sectional dimension,  $N$ , of the panel grow while the time dimension,  $T$ , is kept fixed. We will also assume cross-sectional independence and cross-sectional homoskedasticity throughout the paper.

Ahn and Schmidt (1997) have proposed a GMM estimator for  $r$  in the stationary ARFE(1) model that exhausts all the information in the second moments of the data. One of the moment conditions that are exploited by their GMM estimator is non-linear in  $r$ . By replacing this non-linear moment condition by a combination of moment conditions that are implicit in the Minimum Distance Estimator for the stationary ARFE(1) model (see also Chamberlain, 1982, 1984; Abowd and Card, 1989; and Kruiniger (2000a)), we are able to construct a linear GMM estimator that is asymptotically equivalent to the non-linear GMM estimator of Ahn and Schmidt <sup>1</sup>.

We will also examine the distributional properties of various GMM estimators for both types of models when  $r$  is close to or equal to the unit root. The results we obtain are relevant in many cases, for instance when estimating (dynamic) factor demand models (see Kruiniger (1996)) and production functions (see Griliches and Mairesse (1998), and Blundell and Bond (1999)).

Persistent data raise at least three issues in connection with GMM estimation. First,

---

<sup>1</sup>Kruiniger (2000a) shows that the optimal version of this linear GMM estimator for  $r$  in the stationary ARFE(1) model is asymptotically more efficient than the Conditional Maximum Likelihood estimator and the Optimal Minimum Distance Estimator. This is due to the fact that the former estimator also exploits information in the levels of the data.

many moment conditions do not identify the unit root. For instance, it is well known that the estimator of Arellano and Bond (1991) breaks down when  $r = 1$  because the instruments (lagged levels,  $y_{i,s}$ ,  $s \leq t - 2$ ) are no longer correlated with the first-differences of the regressors ( $\Delta y_{i,t-1}$ ), since  $E(y_{i,s} \Delta y_{i,t-1}) = E(y_{i,s} \varepsilon_{i,t-1}) = 0$  (see Arellano and Bover (1995), and Kruiniger (1996)). Second, some moment conditions are weak in some sense close to the unit root. The finite-sample distributions of the corresponding estimators may differ substantially from the fixed-parameter, first-order asymptotic distributions. Third, some moment conditions are discontinuous at  $r = 1$ : for the inclusive ARFE(1) model it is easily verified that  $E(y_{i,1} \Delta y_{i,t}) = -\frac{r^{t-2} \sigma^2}{r+1}$  when  $|r| < 1$ , and  $E(y_{i,1} \Delta y_{i,t} | y_{i,1}) = 0$  when  $r = 1$ .

In this paper we show that it is important to distinguish between weak moment conditions of the type most often discussed in the literature where the expectation of the product (the covariance) of the instrument and the regressor, or more generally the first-order derivative of the moment condition with respect to the parameter of interest, tends to zero when  $r$  approaches one, and (weak) moment conditions that are characterized by exploding variances (and sometimes also covariances) of the products of the instrument and (some of) the variables from the model. In both cases a form of local asymptotics provides a better approximation to the finite sample distribution of the corresponding GMM estimator than conventional fixed-parameter first-order asymptotics. When the moment conditions are weak in the first (traditional) sense we can obtain approximations to the finite sample distributions of the corresponding GMM estimators by applying local-to-zero asymptotics in the spirit of Staiger and Stock (1997). For the case where the moment conditions are characterized by exploding variances, we propose a local-to-unity asymptotic approach to obtain approximations to the finite sample distributions of the corresponding GMM estimators. This approach amounts to choosing a parameter sequence for which the variances of the cross-products of the instruments and the regressors are  $O(1)$ . We will call such a parameter sequence a local-to-unity parameter sequence.

We find that the first and second moment conditions for the conditional model are all weak in the traditional sense close to the unit root, but that the linear Arellano-Bond moment conditions are no longer weak in that sense in the context of the inclusive model. However the finite sample distributions of the Arellano-Bond and Arellano-Bover estimators for the inclusive model are considerably affected by the fact that the variances (and covariances)

of (one/some of) the products of the instruments and the variables from the model explode when  $r$  approaches one.

Under a local-to-unity parameter sequence the cross-product(s) of the instrument(s) and the regressor(s) in the Arellano-Bond estimator for the inclusive model remain(s) correlated with the cross-product(s) involving the dependent variable when  $N \rightarrow \infty$ , thereby causing (explaining) the bias reported in the Monte Carlo study by Blundell and Bond (1998). Moreover, under the local-to-unity asymptotics the Arellano-Bover estimator has a larger variance than what is predicted by conventional asymptotics, which again is more in agreement with Monte Carlo evidence.

Finally, we also explore the possibility of using GMM estimators for the purpose of testing the unit root hypothesis. To that end we examine the properties of various GMM estimators for both versions of the ARFE(1) model when the true value of  $r$  is equal to one. We find that the two-step Optimal Linear GMM estimator for the inclusive ARFE(1) model is superefficient in this case. This is due to a discontinuity in the derivative of one of the moment conditions. The moment condition in question only identifies the unit root. Another finding is that none of the moment conditions for the conditional ARFE(1) model identify the unit root. Subsequently, we discuss various GMM based tests of the unit root hypothesis against stationary and conditional alternatives respectively.

The paper is organised as follows. Section 2 discusses GMM estimation of the (covariance) stationary ARFE(1) model and presents the Linear GMM estimator for this model. Section 3 reviews GMM estimation of the conditional ARFE(1) model. Section 4 considers estimation of the inclusive ARFE(1) model and the conditional ARFE(1) model when the data are persistent, studies the finite-sample properties of the estimators, and discusses various unit root tests. Section 5 concludes.

## 2 GMM estimation of the stationary ARFE(1) model

In this section and section 3 we consider GMM estimators of parameter  $r$  in two versions of the AR(1) model with Fixed Effects ( ARFE(1))

$$\begin{aligned}
 y_{i,t} &= ry_{i,t-1} + (1-r)\mu_i + \varepsilon_{i,t}, \quad t = t_0 + 1, \dots, T & (1) \\
 \check{\mu}_i &= (1-r)\mu_i \\
 \varepsilon_i | \check{\mu}_i, y_{i,t_0} &\sim (0, \sigma^2 I_{T-t_0}) \quad (\text{ind. dist.}), \\
 E(\varepsilon_{i,t}^4 | \check{\mu}_i, y_{i,t_0}) &< \infty, \quad t = t_0 + 1, \dots, T \\
 t_0 &\leq 1
 \end{aligned}$$

where  $i \in \{1, 2, \dots, N\}$  indicates the individual unit,  $t$  indicates the time period, and  $\varepsilon_i = (\varepsilon_{i,t_0+1}, \dots, \varepsilon_{i,T})$ . For each individual unit we have  $T$  observations on  $y : y_{i,1}, \dots, y_{i,T}$ . The designation “fixed effects” should be taken as to reflect the absence of any distributional assumptions about the individual effects  $\mu_i$ , except that they are independent of the disturbances  $\varepsilon_{i,t}$  given the initial (possibly unobserved) realization of the process,  $y_{i,t_0}$ . In the unit root case the individual effects,  $\mu_i$ , disappear.

The two versions of the ARFE(1) model that we will consider differ with regard to the assumptions that are made about the initial conditions. In section 3 we will review GMM estimation of the ARFE(1) model conditional on the initial observations. However, if the distribution of the initial observations is known, conditioning on these observations when estimating the parameters would entail a loss of information and as a consequence the estimator obtained would (in general) be less efficient than the most efficient estimator. An important case where the distribution of the initial observations is known (up to common parameters) is the situation where the data follow a stationary process. In this section we will discuss GMM estimation of the (covariance) stationary ARFE(1) model.

We will now state the covariance stationarity assumptions

$$\begin{aligned}
 \text{STAT1:} & \quad |r| < 1 \\
 \text{STAT2:} & \quad y_{i,t_0} \sim \left(0, \frac{\sigma^2}{1-r^2}\right)
 \end{aligned}$$

Assumptions STAT1-STAT2 imply that  $\{y_{i,t}\}$  is covariance stationary.

Before we discuss the GMM estimators for the stationary ARFE(1) model, we will introduce more notation. Let  $y_i$  denote the  $T$ -vector of all the observations including the first observation for individual  $i$ . If we denote the covariance matrix of the  $y_i$ 's by  $\sigma^2 V_T(r)$ , we can restate the stationary version of model (1) as

$$(y_i - \mu_i \iota) \equiv u_i \sim (0, \sigma^2 V_T(r)), \quad i = 1, \dots, N, \quad (\text{ind. dist.}) \quad (2)$$

where  $\iota$  is a  $T$ -vector of ones and  $V_T(r)$  has a specific AR(1) structure <sup>2</sup>

$$V_T(r) = \frac{1}{1 - r^2} \begin{pmatrix} 1 & r & r^2 & \dots & r^{T-1} \\ r & 1 & r & \dots & r^{T-2} \\ r^2 & r & 1 & \dots & \dots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ r^{T-1} & r^{T-2} & \dots & r & 1 \end{pmatrix} \quad \text{with } |r| < 1 \quad (3)$$

Kruiniger (2000a) shows that under normality of the  $u_i$ 's the Conditional Maximum Likelihood Estimator for  $r$  and  $\sigma^2$  is consistent and equal to the First Difference (FD) MLE that was proposed by MaCurdy (1981, 1982). The FDMLE corresponds to

$$D(y_i - \mu_i \iota) = Dy_i = Du_i \sim N(0, \sigma^2 DV(r)D'), \quad i = 1, \dots, N, \quad (i.i.d.) \quad (4)$$

where  $D$  denotes the  $(T - 1 \times T)$  first-difference matrix

$$D_T = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots \\ \vdots & 0 & -1 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & \dots & \dots & 0 & -1 & 1 \end{pmatrix}$$

The covariance matrix of  $Dy_i$ ,  $\sigma^2 DV(r)D'$ , has an ARMA(1,1) structure with known MA parameter (-1).

We will now review the moment conditions for the AR(1) model with individual mean effects that can be found in the literature. The most basic moment conditions for this model have been proposed by Arellano and Bond (1991). They follow from the assumption

$$E(\varepsilon_{i,t} | y_{i,t-s}) = 0 \quad \text{for } s = 1, \dots, t - 1 \quad \text{and } t = 2, \dots, T$$

---

<sup>2</sup>We will drop the subscript  $T$  when knowledge of the time-dimension of the panel is unimportant.

We will also omit the argument  $r$  of  $V$ .



This assumption implies  $0.5(T - 1)(T - 2)$  basic moment conditions

$$E(y_{i,t-s}\Delta\varepsilon_{i,t}) = 0 \text{ for } s = 2, \dots, t - 1 \text{ and } t = 3, \dots, T \quad (5)$$

where  $\Delta\varepsilon_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$ . Often one also makes the assumption that the errors are uncorrelated with the individual effects

$$E(\varepsilon_{i,t}\mu_i) = 0 \text{ for } s = 1, \dots, t - 1 \text{ and } t = 2, \dots, T$$

This assumption gives rise to  $T - 3$  additional moment conditions (see Ahn and Schmidt (1995))

$$E((y_{i,T} - ry_{i,T-1})\Delta\varepsilon_{i,t-1}) = 0 \text{ for } t = 4, \dots, T \quad (6)$$

Notice that these moment conditions are non-linear in  $r$ . An alternative assumption is the following

$$E[(y_{i,t} - y_{i,t-1})\check{\mu}_i] = 0 \text{ for } t = 2, \dots, T$$

This assumption adds  $T - 2$  moment conditions to the basic moment conditions, which are also linear in  $r$  (see Arellano and Bover (1995))

$$E((y_{i,t} - ry_{i,t-1})\Delta y_{i,t-1}) = 0 \text{ for } t = 3, \dots, T \quad (7)$$

Using these moment conditions renders (6) redundant for estimation. Finally, adding the assumption of homoskedasticity

$$E(\varepsilon_{i,t}^2) = \sigma^2 \text{ for } t = 2, \dots, T$$

implies another  $T - 2$  moment conditions (see Ahn and Schmidt (1995))<sup>3</sup>

$$E(y_{i,t}\varepsilon_{i,t} - y_{i,t-1}\varepsilon_{i,t-1}) = 0 \text{ for } t = 3, \dots, T \quad (8)$$

Notice that these moment conditions are also linear in  $r$ . Therefore, when all the assumptions listed above hold, a GMM estimator for  $r$  can be constructed that exploits a total of  $0.5T(T + 1) - 3$  linear moment conditions.

---

<sup>3</sup>In fact we only need homoskedasticity in the time dimension  $E(\varepsilon_{i,t}^2) = \sigma_i^2$  for  $t = 2, \dots, T$  and  $\forall i$ .

All the moment conditions (and assumptions) just mentioned are compatible with the stationary ARFE(1) model that was described above. In fact all the moment conditions given in (5)-(8) still hold for a weaker version of the stationary ARFE(1) model that assumes  $\varepsilon_i \sim (0, \sigma_i^2 I_{T-1})$  (*ind. dist.*) ;  $E(\check{\mu}_i \varepsilon_i) = 0$  ;  $E(y_{i,1} \varepsilon_i) = 0$ , and  $\{y_{i,t}\}$  is covariance stationary. However, we have not yet exhausted the set of all the moment conditions that are implied by covariance stationarity of  $\{y_{i,t}\}$ . Covariance stationarity of  $\{y_{i,t}\}$  implies a total of  $0.5T(T+1)$  moment conditions corresponding to  $E(\text{vech}(y_i y_i'))$ , or equivalently  $0.5T(T+1)$  moment conditions corresponding to  $E[\text{vech}((Dy_i)(Dy_i)')]$ ,  $E(Dy_i y_{i,1})$ , and  $E(y_{i,1}^2)$  <sup>4</sup>. Clearly, these moment conditions not only involve  $r$  but also  $\sigma^2$ . Moreover, when  $|r| < 1$ , the last moment condition also depends on  $\mu_i^2$ . When  $r = 1$ , estimation proceeds by conditioning on  $y_{i,1}$  and the last moment condition disappears altogether. Therefore the last moment condition is not instrumental for estimating  $r$ . We lose another moment condition when eliminating  $\sigma^2$  ( $\sigma_i^2$ ). Ahn and Schmidt (1997) found that covariance stationarity of  $\{y_{i,t}\}$  implies the missing moment condition <sup>5</sup>

$$E(y_{i,1}^2 + y_{i,2} \Delta \varepsilon_{i,3} / (1 - r^2) - (y_{i,3} - r y_{i,2})(y_{i,2} - r y_{i,1}) / (1 - r)^2) = 0 \quad (9)$$

Notice that this moment condition is non-linear in  $r$ . Therefore computing the value of the GMM estimator of Ahn and Schmidt (1997) that exploits all the moment conditions that are implied by covariance stationarity, (5)-(9), requires numerical optimization.

## 2.1 An efficient linear GMM estimator for the stationary ARFE(1) model

In this subsection we will propose a new, fully efficient, linear GMM estimator for the stationary model that is asymptotically equivalent to the nonlinear Optimal GMM estimator of Ahn and Schmidt (1997).

---

<sup>4</sup>Covariance stationarity also implies restrictions on the first moments:  $E(\Delta y_{i,t}) = 0$ ,  $t = 2, \dots, T$ .

Crépon et al. (1997) discusses first moment conditions for the AR(1) model with random effects.

<sup>5</sup>Ahn and Schmidt have not shown that this moment condition is not redundant relative to the  $0.5T(T+1) - 3$  linear moment conditions from the literature (5), (7) and (8).

Kruiniger (2000a) shows that under normality the Conditional MLE (First Difference MLE) is asymptotically equivalent to the Optimal GMM estimator which exploits the moment conditions in

$$E(\text{vech}((Dy_i)(Dy_i)' - \Omega(r, \sigma^2))) = 0 \quad (10)$$

where  $\Omega(r, \sigma^2) = \sigma^2 DV(r)D'$ . These moment conditions summarize all the information in the second moments of the (first-)differenced data.

It turns out that the moment conditions in (10) imply a moment condition that is very useful for the estimation of  $r$ . We have the following result

**Theorem 1** *Under normality the linear moment condition*

$$E((r - 1)(\Delta y_{i,2})^2 - 2\Delta y_{i,2}\Delta y_{i,3}) = 0 \quad (11)$$

*is not redundant relative to the linear moment conditions from the literature, i.e. (5), (7) and (8), when  $|r| < 1$  and  $T = 3$ .*

PROOF

See appendix A. It is also shown that (8) is not redundant relative to (5) and (7), and that (7) is not redundant relative to (5). The property of redundancy is defined by Breusch et al. (1999). The case  $r = 1$  will be discussed in section 4.

We define the Linear GMM (LGMM) estimator as the estimator that exploits the  $0.5T(T + 1) - 3$  linear moment conditions from the literature, (5), (7) and (8), and the new linear moment condition, (11). The Linear GMM estimator exhausts the set of all the restrictions on the second moments that are implied by the assumption of covariance stationarity of  $\{y_{i,t}\}$ . It follows (see also appendix A) that the two-step Optimal Linear GMM (OLGMM) estimator is asymptotically equivalent to the non-linear Optimal GMM estimator due to Ahn and Schmidt (1997) that exploits the non-linear moment condition (9) instead of (11) but otherwise the same moment conditions as the (Optimal) LGMM estimator. Of course, the OLGMM estimator is preferable to the non-linear GMM estimator, as it is more easily calculated. Moreover its properties of the LGMM estimator are more easily investigated. In section 4 we will see that the LGMM estimator is also consistent in the unit root case.

### 3 GMM estimation of the conditional ARFE(1) model

Sometimes we cannot make the stationarity assumption. For instance, when the initial observations  $y_{i,1}$ ,  $i = 1, 2, \dots, N$ , are start-up values of stochastic processes, they may follow a distribution that is different from the distribution of the subsequent observations. In that case we may decide to condition on the initial observations, i.e. to treat the initial observations as given. The conditional AR(1) model with fixed effects can be stated as

$$\begin{aligned}
 y_i &= ry_{i,-1} + \check{\mu}_{it} + \varepsilon_i & (12) \\
 \check{\mu}_i &= (1 - r)\mu_i \\
 |y_{i,1}| &< \infty \\
 \varepsilon_i | \check{\mu}_i, y_{i,1} &\sim (0, \sigma^2 I_{T-1}) \text{ (ind. dist.)}, \\
 E(\varepsilon_{i,t}^4 | \check{\mu}_i, y_{i,1}) &< \infty, \quad t = 2, \dots, T
 \end{aligned}$$

where  $i \in \{1, 2, \dots, N\}$ . For each individual unit, we have  $T$  observations on  $y$  including the initial observation  $y_{i,1}$ . In the unit root case the individual effects,  $\mu_i$ , disappear.

A weaker version of the conditional ARFE(1) model assumes

$$\varepsilon_i \sim (0, \sigma_i^2 I_{T-1}) \text{ (ind. dist.)}; \quad E(\check{\mu}_i \varepsilon_i) = 0; \quad |y_{i,1}| < \infty; \quad E(y_{i,1} \varepsilon_i) = 0 \quad (13)$$

The conditional ARFE(1) model implies the following restrictions on the second moments (cf Ahn and Schmidt, 1997)<sup>6</sup>:

$$\begin{aligned}
 E(y_{i,1} \Delta \varepsilon_{i,t}) &= 0, \quad t = 3, \dots, T, & (14) \\
 E((y_{i,t} - ry_{i,t-1})^2 - (y_{i,2} - ry_{i,1})^2) &= 0, \quad t = 3, \dots, T, \quad \text{and} \\
 E((y_{i,t} - ry_{i,t-1})(y_{i,s} - ry_{i,s-1}) - (y_{i,3} - ry_{i,2})(y_{i,2} - ry_{i,1})) &= 0, \quad 1 \leq s < t, \quad t = 4, \dots, T
 \end{aligned}$$

The moment conditions in (14) exhaust all the restrictions on the second moments that are implied by the conditional ARFE(1) model.

We define the Conditional GMM (CGMM) estimator as the GMM estimator that exploits the moment conditions in (14).

---

<sup>6</sup>We notice that assumptions (A.1)-(A.4) in Ahn and Schmidt (1997) are satisfied by the weaker version of the conditional ARFE(1) model.

In addition to the moment conditions in (14), the conditional ARFE(1) model implies the following restrictions on the first moments

$$E(\Delta y_{i,t} - r\Delta y_{i,t-1}) = 0, \quad t = 3, \dots, T \quad (15)$$

## 4 GMM estimation of ARFE(1) models with persistent data

Persistent data raise at least three problems in connection with GMM estimation. First, many moment conditions do not identify the unit root. For instance, it is well known that the pure Arellano-Bond estimator breaks down when  $r = 1$  because the instruments (lagged levels,  $y_{i,s}$ ,  $s \leq t - 2$ ) are no longer correlated with the first-differences of the RHS variables ( $\Delta y_{i,t-1}$ ), since  $E(y_{i,s}\Delta y_{i,t-1}) = E(y_{i,s}\varepsilon_{i,t-1}) = 0$ , see Arellano and Bover (1995), and Kruiniger (1996). Second, moment conditions may become weak in some sense close to the unit root. Third, some moment conditions are discontinuous at  $r = 1$ . For the inclusive ARFE(1) model it is easily verified that  $E(y_{i,1}\Delta y_{i,t}) = -\frac{r^t - 2\sigma^2}{r+1}$  when  $|r| < 1$  and  $E(y_{i,1}\Delta y_{i,t}|y_{i,t_0}) = 0$  when  $r = 1$ . This implies that  $E(Dy_{i,1})$  is left-discontinuous at  $r = 1$ . Below we will investigate the properties of the moment conditions and the corresponding estimators for both versions of the ARFE(1) model when  $r = 1$  or close to the unit root. In particular, we will distinguish between various (weak) moment condition problems and derive the finite sample distribution of various estimators using local-to-unity asymptotics. In addition, we will discuss various GMM based unit root tests and compare them with other unit root tests from the literature.

### 4.1 GMM estimation of the inclusive ARFE(1) model

In this subsection we will discuss GMM estimation of the inclusive ARFE(1) model, which contains the stationary ARFE(1) models and the ARFE(1) model with an autoregressive unit root, when the data are persistent, and we will discuss various tests of the unit root hypothesis versus stationarity.

We will first give a formal definition of the inclusive ARFE(1) model.

**Definition 1** *The inclusive ARFE(1) model is characterized by (1) and the following assumptions:*

$$\text{INCL1:} \quad -1 < r \leq 1$$

$$\text{INCL2:} \quad y_{i,t_0} \sim (0, \frac{\sigma^2}{1-r^2}) \text{ if } -1 < r < 1; |y_{i,t_0}| < \infty \text{ if } r = 1$$

The unit root case is special because no matter how long ago the process  $\{y_{i,t}\}$  has started, the influence of early realisations of  $y$  and  $\varepsilon$  on the current observation  $y_{i,t}$  does not vanish <sup>7</sup>. As a consequence, when  $r = 1$ , we cannot avoid conditioning on the initial observations. Moreover, we cannot rewrite model (1) as (2) and the matrix  $V$  no longer exists when  $r = 1$ . However, Kruiniger (2000a) shows that  $E(Dy_i y_i' D')$  still equals  $\sigma^2 D V D'$  when  $r = 1$  and is (left-)continuous in  $r = 1$ . <sup>8</sup>

In the sequel we will assume that  $T = 3$ . Then the inclusive ARFE(1) model implies a total of four moment conditions which depend only on  $r$ :

$$E(m_1(r)) = E(\varepsilon_{i,3} \Delta y_{i,2}) = E((y_{i,3} - r y_{i,2}) \Delta y_{i,2}) = 0 \quad (16)$$

$$E(m_2(r)) = E(y_{i,1} \Delta \varepsilon_{i,3}) = E(y_{i,1} (\Delta y_{i,3} - r \Delta y_{i,2})) = 0$$

$$E(m_3(r)) = E(y_{i,3} \varepsilon_{i,3} - y_{i,2} \varepsilon_{i,2}) = E(y_{i,3} (y_{i,3} - r y_{i,2}) - y_{i,2} (y_{i,2} - r y_{i,1})) = 0$$

$$E(m_4(r)) = E((r-1)(\Delta y_{i,2})^2 - 2 \Delta y_{i,2} \Delta y_{i,3}) = 0$$

Notice that these moment conditions are still valid when  $r = 1$  and continuous in  $r$  (also at  $r = 1$ ). Let  $m(r) = (m_1(r) \ m_2(r) \ m_3(r) \ m_4(r))'$ . Evaluating  $m(r)$  at  $r = 1$ , we obtain

$$m_1(1) = \varepsilon_{i,3} \varepsilon_{i,2} \quad (17)$$

$$m_2(1) = y_{i,1} \Delta \varepsilon_{i,3}$$

$$m_3(1) = (\varepsilon_{i,3})^2 + \varepsilon_{i,2} \varepsilon_{i,3} - (\varepsilon_{i,2})^2 + y_{i,1} \Delta \varepsilon_{i,3}$$

$$m_4(1) = -2 \varepsilon_{i,2} \varepsilon_{i,3}$$

---

<sup>7</sup>Another consequence of  $r = 1$  is of course that the incidental parameters  $\mu_i$  drop out from the model, so in this case there is no need for removing them.

<sup>8</sup>Also the conditional likelihood function under normality is (twice) differentiable at  $r = 1$ . Therefore we can still use the CML estimator for the ARFE(1) model that includes both the stationary versions and the unit root version (see Kruiniger (2000a)).

It is easily seen from (17) that when  $r = 1$  the covariance matrix of  $m_1$  and  $m_4$  is singular. Let us now consider  $m_5(r) = 2m_1(r) + m_4(r) = (1 - r)((\Delta y_{i,2})^2 + 2y_{i,1}\Delta y_{i,2})$ . Note that  $m_5(1) = 0$ . Therefore, when  $r = 1$  the optimal GMM estimator is the estimator that gives all weight to  $E(m_5(r)) = 0$ , that is the estimator that only exploits  $E(m_5(r)) = 0$ . This estimator equals 1 with probability one and its variance equals zero. However, this estimator is inconsistent when  $|r| < 1$ . Therefore, in practice one should not use the optimal weighting matrix for  $r = 1$  unless one would know that  $r = 1$  but in that case of course there is no need to estimate  $r$ . Nevertheless, since  $E((\Delta y_{i,2})^2 + 2y_{i,1}\Delta y_{i,2}) = 0$  when  $|r| < 1$ , GMM estimators that combine  $E(m_5(r)) = 0$  with (some) other, independent moment condition(s) contained in  $E(m(r)) = 0$  are consistent for both  $|r| < 1$  and  $r = 1$ .

It should be clear from the discussion above that one can construct a consistent GMM estimator which has arbitrarily small variance when  $r = 1$  by choosing the weight for  $E(m_5(r)) = 0$  sufficiently high relative to the weight given to other moment conditions. In particular, there exist GMM estimators that have a variance that is smaller than the Cramér-Rao lowerbound when  $r = 1$ <sup>9</sup>. The reason for this superefficiency is that  $E(\frac{dm_5(r)}{dr}) = -E((\Delta y_{i,2})^2 + 2y_{i,1}\Delta y_{i,2}) = 0$  when  $|r| < 1$ , whereas  $E((\Delta y_{i,2})^2 + 2y_{i,1}\Delta y_{i,2}) = \sigma^2$  when  $r = 1$ .

Let  $\tilde{m}(r) = (m_2(r) \ m_3(r) \ m_4(r) \ m_5(r))'$ . We note that  $E(\tilde{m}(r)) = 0$  is equivalent to  $E(m(r)) = 0$ . In appendix B.1.1 it is shown that a properly normalised asymptotically optimal weighting matrix for a GMM estimator that exploits  $E(\tilde{m}(r)) = 0$ , e.g. with trace equal to 1, is left-continuous at  $r = 1$ . Therefore a two-step GMM estimator that makes use of this weighting matrix evaluated at the value of an initial consistent estimator will be asymptotically equivalent to the optimal GMM estimator even when  $r = 1$ . We remark that when  $r$  is close to 1 this weighting matrix is ill-conditioned.

#### 4.1.1 Identification and redundancy when $r=1$

We have seen that the Arellano-Bond moment conditions given in (5) do not identify the unit root. In the context of the inclusive ARFE(1) model they are the exception to the rule:

---

<sup>9</sup>The Cramér-Rao lowerbound is given by  $CR^C(1, T) = 2\sigma^2 / [(T - 1)((T - 2)\sigma^2 + 2(p \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N y_{i,1}^2))]$  assuming that  $p \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 < \infty$  (see Krueger (2000a)). Notice the dependence of the Cramér-Rao lowerbound on the initial conditions.

**Theorem 2** *In the inclusive ARFE(1) model, all the moment conditions that are exploited by the Linear GMM estimator identify the unit root except for the basic moment conditions of Arellano and Bond (1991) given in (5).*

PROOF

See appendix B.1.1.

From now on we assume that the  $y_{i,t}$  ( $\varepsilon_{i,t}$ ) are normally distributed unless we state otherwise.

In order to investigate whether the Arellano-Bond moment condition is actually also redundant when  $r = 1$ , we will compare the asymptotic variances of various Optimal GMM estimators that exploit subsets of the four moment conditions in  $E(m(r)) = 0$  which do not include both the first and the last moment condition.

The asymptotic variance of the GMM estimator that exploits the last three moment conditions in  $E(m(r)) = 0$  is equal to  $\frac{16}{5}$  when  $r = 1$ . This GMM estimator is more efficient than the GMM estimator that leaves out  $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$ , but still exploits the third and fourth moment conditions. This efficiency gain results from the fact that the second moment condition is correlated with the third and the fourth moment conditions in  $E(m(r)) = 0$ . One can therefore combine  $m_3(r)$  and  $m_4(r)$  with  $m_2(r)$  in such a way that the difference between the covariance matrix of the new moment conditions and the covariance matrix of  $m_3(r)$  and  $m_4(r)$  is negative definite <sup>10</sup>.

The GMM estimator that only exploits  $E((r-1)(\Delta y_{i,2})^2 - 2\Delta y_{i,2}\Delta y_{i,3}) = 0$  is asymptotically equivalent to the GMM estimator for  $r$  which is based on the three moment conditions in  $E(\text{vech}(Z_i)) - \text{vech}(\Omega(r, \sigma^2)) = 0$ : their asymptotic variance equals 4 when  $r = 1$ .

In Kruiniger (2000a) it was shown that under normality,  $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$  is not redundant relative to the set of moment conditions  $E(\text{vech}(Z_i)) - \text{vech}(\Omega(r, \sigma^2)) = 0$  when  $|r| < 1$  and  $T = 3$ . In the unit root case we have a different result:

---

<sup>10</sup>This can easily be seen as follows. Let  $g_1 = m_3(r)$ ,  $g_2 = m_2(r)$ ,  $g = (g_1 \ g_2)'$  and  $M = E(gg')$ . Next define  $h_1 = g_1 - M_{12}M_{22}^{-1}g_2$ . Then observe that the estimator that only minimizes  $(\sum_i h_{1,i})^2$  is more efficient than the estimator that minimizes  $(\sum_i g_{1,i})^2$ , since  $M_{12}^2 \neq 0$  and therefore  $\text{Var}(h_{1,i}) = M_{11} - M_{12}^2M_{22}^{-1} < M_{11} = \text{Var}(g_{1,i})$ .



**Theorem 3** *When  $r = 1$  and the  $\varepsilon_{i,t}$  are symmetrically distributed, the moment conditions  $E(y_{i,1}\Delta\varepsilon_{i,t}) = 0$  for  $t = 3, \dots, T$  are redundant relative to  $E(\text{vech}(Z_i)) - \text{vech}(\Omega(r, \sigma^2)) = 0$ .*

PROOF

When  $r = 1$ ,  $E(\frac{d}{dr}(y_{i,1} \times (\Delta y_{i,t} - r\Delta y_{i,t-1}))) = -E(y_{i,1}\varepsilon_{i,t-1}) = 0$ ,  $t = 3, \dots, T$ . The theorem follows straightforwardly from the fact that the optimal weighting matrix is diagonal.  $\square$

The asymptotic variance of the GMM estimator that exploits the first three moment conditions in  $E(m(r)) = 0$  is equal to 1, which is also the asymptotic variance of the GMM estimator that only exploits the Arellano-Bover moment condition  $E(\varepsilon_{i,3}\Delta y_{i,2}) = 0$ . This implies that the second and third moment condition in  $E(m(r)) = 0$  are redundant relative to  $E(\varepsilon_{i,3}\Delta y_{i,2}) = 0$ . However, in practice it may be better to use for instance the first two moment conditions, i.e. the System estimator (see e.g. Blundell and Bond (1998)), since this would result in more efficient estimates asymptotically when  $r < 1$ .

#### 4.1.2 Weak moment conditions and local-to-unity asymptotics

It can easily be shown that the (fixed-parameter) asymptotic variances of the Optimal GMM estimators which are based on the first three, on the last three, and on all four moment conditions in  $E(m(r)) = 0$ , respectively, all tend to 4 when  $r \uparrow 1$ <sup>11</sup>. It follows that the asymptotic variances of the Optimal GMM estimators that exploit either the first three moment conditions in  $E(m(r)) = 0$  or the last three are discontinuous at  $r = 1$ . This is due to the fact that the first three elements of  $E(\frac{dm}{dr})$  are discontinuous at  $r = 1$ <sup>12</sup>. For instance,  $E(\frac{d}{dr}(y_{i,1}(\Delta y_{i,3} - r\Delta y_{i,2}))) = -E(y_{i,1}\Delta y_{i,2})$  is discontinuous at  $r = 1$  since  $\lim_{r \uparrow 1} E(y_{i,1}\Delta y_{i,2}) = -\frac{1}{2}\sigma^2$ , whereas  $E(y_{i,1}\Delta y_{i,2}) = E(y_{i,1}\varepsilon_{i,2}) = 0$  when  $r = 1$ . Thus under

---

<sup>11</sup>The limits of these fixed parameter asymptotic variances are only calculated to make a theoretical point. In practice, when  $r$  is close to 1, one should rely on local-to-unity asymptotic results (variances) that are discussed below.

<sup>12</sup>See also appendix B. Above we have seen that for the ARFE(1) model  $E(y_{i,1}\Delta y_{i,t})$ ,  $t = 2, \dots, T$ , are discontinuous at  $r = 1$ . It follows that for any  $T \geq 3$  the derivatives w.r.t.  $r$  of all the moment conditions that are exploited by the Linear GMM estimator, apart from the new linear moment condition (11), are discontinuous at  $r = 1$ .

stationarity the Arellano-Bond GMM estimator does *not* suffer from a traditional weak instruments (moments) problem where the expectation of the product (the covariance) of the instrument and the regressor would tend to zero when  $r$  approaches one<sup>13</sup> (see also Nelson and Startz (1990a, b), Maddala and Jeong (1992), Bound, Jaeger, and Baker (1995), and Staiger and Stock (1997)). Nevertheless, as Blundell and Bond (1998) show, in the case of the Arellano-Bond estimator, the concentration parameter of the distribution of the first-stage F-statistic divided by  $N$ , and the slope parameters in the first-stage regression tend to zero when  $r$  approaches one. Moreover, Kruiniger (1996) observed that  $\lim_{r \uparrow 1} Avar(\hat{r}_{GMMAB}) = \infty$ .

With regard to the weak instruments problem we have the following result:

**Theorem 4** *In the inclusive ARFE(1) model, none of the moment conditions that are exploited by the Linear GMM estimator are weak in the (traditional) sense that the first-order derivative of the moment condition with respect to  $r$  tends to zero when  $r$  approaches one except for the ‘homoskedasticity moment conditions’ of Ahn and Schmidt (1997) given in (8).*

#### PROOF

See appendix B.1.1. The non-linear moment conditions in (14) are still valid for the stationary model and indeed they follow from the moment conditions exploited by the Linear GMM estimator. In appendix B.2.1. we show that these non-linear moment conditions are weak under both versions of the model.

Notwithstanding that the Arellano-Bond and the Arellano-Bover moment conditions are not weak in the traditional sense, the finite sample distributions of the corresponding estimators differ considerably from the fixed-parameter first-order asymptotic distributions when  $r$  is close to one (see for instance the Monte Carlo evidence in Blundell and Bond (1998)). In particular, the finite sample distributions are affected by exploding variances (and covariances) of (some of) the cross-products of the instruments and the variables from the model when  $r$  approaches one. For both estimators local-to-unity asymptotics provide better approximations to their finite sample distributions than conventional fixed-parameter

---

<sup>13</sup>However, in the conditional model the Arellano-Bond moment condition  $E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$  is a weak moment condition in this sense when  $r$  is close to 1.

asymptotics. We obtain the local-to-unity approximations to the distributions of the corresponding estimators by choosing a parameter sequence such that the variances of the cross-products of the instruments and the regressors become  $O(1)$ . For  $T = 3$  we have the following local-to-unity asymptotic results for the Arrelano-Bond and Arrelano-Bover estimators:

**Theorem 5** *Let  $r = 1 - \frac{\lambda}{N}$ , and  $T = 3$ . Then  $\text{plim}_{N \rightarrow \infty} \hat{r}_{ARBOV} = r$ ,*

$$\hat{r}_{ARBOND} \xrightarrow{d} r + \frac{\tilde{X}_1}{\tilde{X}_2}, \quad \text{where} \quad \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix} \sim N \left[ -\frac{\sigma^2}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \frac{\sigma^4}{2\lambda} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right], \quad \text{and}$$

$$\sqrt{N}(\hat{r}_{ARBOV} - r) \xrightarrow{d} \frac{\tilde{X}_3}{\tilde{X}_4}, \quad \text{where} \quad \begin{bmatrix} \tilde{X}_3 \\ \tilde{X}_4 \end{bmatrix} \sim N \left[ \frac{\sigma^2}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \sigma^4 & 0 \\ 0 & \frac{\sigma^4}{2\lambda} \end{pmatrix} \right].$$

PROOF

See appendix B.1.2. There we also derive alternative local-to-unity asymptotic results corresponding to the model  $y_{i,t} = ry_{i,t-1} + \eta_i + \varepsilon_{i,t}$  where we keep  $\eta_i (= (1-r)\mu_i)$  fixed instead of  $\mu_i$ . In order to obtain these alternative, fixed  $\eta_i$  local-to-unity asymptotic results, we use the parameter sequence  $r = 1 - (\frac{\lambda}{N})^{0.5}$  instead of  $r = 1 - \frac{\lambda}{N}$ . This implies that when  $\eta_i$  is kept fixed, the finite sample phenomena disappear more slowly when  $N$  increases than in the fixed  $\mu_i$  case.

The distribution of the ratio of two, possibly correlated, normal variables has already been studied by Fieller (1932). This ratio does not have finite moments.

We also have local-to-unity asymptotic results for  $T > 3$ .

**Theorem 6** *Let  $r = 1 - \frac{\lambda}{N}$ , and  $T > 3$ . In addition, let  $W_N$  be an arbitrary sequence of PDS weighting matrices, with  $\text{plim}_{N \rightarrow \infty} W_N = W$ , where  $W$  is PDS. Then  $\text{plim}_{N \rightarrow \infty} \hat{r}_{ARBOV} = r$ ,*

$$\hat{r}_{ARBOND} \xrightarrow{d} r + \frac{\tilde{X}'_6 W \tilde{X}_5}{\tilde{X}'_6 W \tilde{X}_6}, \quad \text{where } \tilde{X}_5, \tilde{X}_6 \text{ are normally distributed, and}$$

$$\sqrt{N}(\hat{r}_{ARBOV} - r) \xrightarrow{d} \frac{\tilde{X}'_8 W \tilde{X}_7}{\tilde{X}'_8 W \tilde{X}_8}, \quad \text{where } \tilde{X}_7, \tilde{X}_8 \text{ are also normally distributed.}$$

Moreover  $\hat{r}_{ARBOND}$  is asymptotically biased downwards, whereas  $\hat{r}_{ARBOV}$  is asymptotically unbiased.

PROOF

See appendix B.1.3. We will prove that  $\widehat{r}_{ARBOND}$  is asymptotically biased downwards for the more interesting case where the optimal weighting matrix is used. The other results can be proven in a similar way.

**Corollary 7** *For  $T > 3$  the optimal Arellano-Bond estimator is asymptotically biased downwards when  $r = 1 - \frac{\lambda}{N}$ .*

When considering the parameter sequence  $r = 1 - \frac{\lambda}{N}$ , the cross-products of the instruments and the regressors (when  $T = 3$  the denominator) in the Arellano-Bond estimator for the inclusive model remain(s) correlated with the cross-products involving the dependent variables (the numerator when  $T = 3$ ) when  $N \rightarrow \infty$ , thereby causing (explaining) the bias reported in the Monte Carlo study by Blundell and Bond (1998). Moreover, under these local-to-unity asymptotics the Arellano-Bover estimator has a larger variance than what is predicted by conventional asymptotics, which again is more in agreement with Monte Carlo evidence, but is still asymptotically (median) unbiased. However, in the alternative, fixed  $\eta_i$  framework the Arellano-Bover estimator is found to be asymptotically biased too.

In the optimal System GMM estimator, which exploits both the Arellano-Bond and the Arellano-Bover moment conditions, the Arellano-Bond moment condition receives less and less weight and becomes redundant when  $r \uparrow 1$  (see above and appendix B). Intuitively, this makes sense because the local-to-unity asymptotic results indicate that the variance of the Arellano-Bover estimator decreases at rate  $N$  whereas the variance of the Arellano-Bond estimator does not change when  $N$  increases. We conclude that the local-to-unity asymptotic results obtained for the Arellano-Bover estimator also bear on the optimal System estimator.

**Theorem 8** *The optimal System GMM estimator, which exploits both the Arellano-Bond and the Arellano-Bover moment conditions, has the same local-to-unity asymptotic properties as the Arellano-Bover estimator, when choosing the parameter sequence  $r = 1 - \frac{\lambda}{N}$ .*

#### PROOF

See appendix B.1.4. One can also show along the lines of this proof and using the results in appendix B.1.1 that local-to-unity the optimal Linear GMM estimator — which exploits  $E(\tilde{m}(r)) = 0$  — converges in distribution at rate  $N$  (instead of  $N^{0.5}$ ).

### 4.1.3 Unit root tests

One can exploit the discontinuity in  $E(\frac{dm_5(r)}{dr})$  at  $r = 1$  to construct powerful tests of  $r = 1$  versus stationarity<sup>14</sup>. As an example, consider a simple test that is based on a GMM estimator that exploits  $E(m_4(r)) = 0$  and  $E(m_5(r)) = 0$ , using a diagonal weighting matrix with weights  $(1 - w)$  and  $w$  respectively, where  $0 < w < 1$ . This estimator is consistent both under  $r = 1$  and under stationarity. Assuming normality the asymptotic variance of this GMM estimator equals  $4(1 - w)^2$  under the null hypothesis, while under the alternative hypothesis its asymptotic variance equals  $4(11 - 10r + 3r^2)(1 + r)^{-2}$  which is larger than  $4 \forall |r| < 1$  and also independent of  $w$ <sup>15</sup>. So the width of the confidence interval depends on  $w$ : the higher the value of  $w$ , the more powerful the test is asymptotically.

We will now examine the properties of pure Arellano-Bover type GMM estimators when  $r = 1$ . We will drop the normality assumption and just assume  $\varepsilon_{i,t} \sim i.i.d.(0, \sigma^2)$  and finite fourth moments. Then we have the following results:

**Theorem 9** *When  $T = 3$ ,  $t_0$  is fixed and  $|y_{i,t_0}| < \infty$ , the GMM estimator that is only based on  $E((y_{i,3} - ry_{i,2})\Delta y_{i,2}) = 0$  is consistent when  $r = 1$ .*

PROOF

Trivial. See appendix B.1.5.

**Theorem 10** *When  $t_0$  and  $T$  are fixed,  $|y_{i,t_0}| < \infty$ , and  $r = 1$ , the asymptotic covariance of the Optimal GMM estimator that is based on the moment conditions  $E((y_{i,t} - ry_{i,t-1})\Delta y_{i,t-s}) = 0$  for  $t = 3, \dots, T$  and  $s \geq 1$  is equal to  $\frac{2}{(T-1)(T-2)}$ .*

---

<sup>14</sup>When  $r = 1$  one extra moment condition is available,  $E(y_{i,1}(y_{i,2} - ry_{i,1})) = 0$ , which is invalid when  $|r| < 1$ . However, in this subsection we restrict our attention to test statistics that are based on GMM estimators which are consistent both when  $r = 1$  and when  $|r| < 1$ .

<sup>15</sup>When  $r$  is close to one, a better approximation to the distribution of this estimator can be obtained by employing local-to-unity asymptotics which is discussed later in this subsection (see also appendix B.1, especially B.1.4.)

PROOF

Since  $\varepsilon_{i,t} \sim i.i.d.(0, \sigma^2)$ , the optimal weighting matrix is equal to  $\sigma^{-4}I$ . Moreover,  $E(\frac{d}{dr}(y_{i,t} - ry_{i,t-1})\Delta y_{i,t-s}) = -E(y_{i,t-1}\Delta y_{i,t-s}) = -E(y_{i,t-1}\varepsilon_{i,t-s}) = -\sigma^2$ .  $\square$

It follows that under normality the asymptotic variance of the Optimal Arellano-Bover type GMM estimator is four times smaller than the asymptotic variance of the CMLE, (which is equal to the asymptotic variance of the Optimal GMM estimator that is based on  $E(\text{vech}(Z_i)) - \text{vech}(\Omega(\theta)) = 0$ ) when  $r = 1$ .

We can easily show that the weighting matrix of the Optimal Arellano-Bover GMM estimator converges to  $\sigma^{-4}I$  when  $r \uparrow 1$  and that the asymptotic variance of this estimator converges to  $\frac{8}{(T-1)(T-2)}$  when  $r \uparrow 1$ . We conclude that a unit root test based on the Arellano-Bover moment conditions does not require any distributional assumptions to be satisfied and is under normality asymptotically more powerful against local alternatives than the CMLE-based test. An additional advantage of this kind of unit root tests is that they can accommodate heterogeneity in the MA parameters.

Finally, the GMM estimator that exploits  $E(m_5(r)) = 0$  and an optimal combination of the Arellano-Bover moment conditions  $E(\varepsilon_{i,t}\Delta y_{i,t-s}) = 0$ ,  $s = 1, \dots, t-2$ ,  $t = 3, \dots, T$ , and uses  $w$  and  $(1-w)$  as weights for the former and the latter, respectively, is consistent for  $-1 < r \leq 1$ . When  $r = 1$  the asymptotic variance of this estimator equals  $(\frac{1}{2}(T-1)(T-2) \times (1-w) + w)^{-2} (\frac{1}{2}(T-1)(T-2)(1-w)^2)$ . Notice that the asymptotic variance is decreasing in  $w$ . So the estimator leads to a very simple but relatively speaking quite powerful test of  $r = 1$ . The optimal choice of  $w$  will depend on the dimensions of the panel data set.

Various other unit root tests for panels with fixed  $T$  have already been suggested in the literature. Breitung and Meyer (1994) have proposed an estimator for the inclusive ARFE(1) model which can be used as a test statistic. This estimator is based on deviations from the initial observations. It is consistent both when  $r = 1$  and when  $|r| < 1$  (just like the CMLE), and has the same large  $N$  asymptotic variance as the CMLE. Harris and Tzavalis (1999) present a unit root test statistic for the conditional ARFE(1) model which is an adjusted version of the Least Squares Dummy Variables estimator. The adjustment corrects for the inconsistency of the LSDV estimator. The corrected estimator has asymptotic variance  $10.2/(T-1)(T-2)$  (approximately) for fixed (but large)  $T$  and  $N \rightarrow \infty$ .

## 4.2 GMM estimation of the conditional ARFE(1) model with persistent data

In this subsection we will study the properties of the moment conditions for the conditional ARFE(1) model when  $r$  is close to or equal to the unit root. In particular, we will address the issue of weak moment conditions in the context of the conditional ARFE(1) model, and we will discuss the unit root case.

With regard to weak moment conditions we have the following result

**Theorem 11** *All the moment conditions corresponding to the conditional ARFE(1) model, i.e. the moment conditions in (14) and (15), are weak in the traditional sense when  $r$  is close to unity.*

PROOF

Trivial, see Appendix B.2.1.

In section 4.1.2, we have seen that the Arellano-Bond moment conditions given in (5) are *not* weak in the traditional sense for the inclusive ARFE(1) model when  $r$  is close to one. Furthermore, some moment conditions for the inclusive model remain informative when  $r$  approaches one, whereas theorem 11 implies that the moment conditions for the conditional ARFE(1) model contain virtually no information when  $r$  is close to one. This is in agreement with the findings of Hahn (1999), who showed that the ratio of the semi-parametric information bounds for the conditional ARFE(1) model and for the stationary ARFE(1) model, tends to infinity when  $r$  approaches one, implying that the information contained in the conditional model with regard to  $r$  decreases to zero when  $r$  approaches one.

Staiger and Stock (1997) argued that local-to-zero asymptotics provide a better approximation to the finite sample distribution of a GMM estimator that exploits moment conditions that are weak in the traditional sense. Local-to-zero refers to the local-to-zero covariance between the weak instrument and the regressor. The approach of Staiger and Stock yields the following approximation to the finite sample distribution of the Arellano-Bond GMM estimators for the conditional ARFE(1) model when  $r$  is close to one:

**Theorem 12** Let  $r = 1 - \lambda N^{-0.5}$ , and  $T = 3$ . Then  $\hat{r}_{ARBOND} \xrightarrow{a} r + \frac{\check{Z}_1}{\check{Z}_2}$ , where

$$\begin{bmatrix} \check{Z}_1 \\ \check{Z}_2 \end{bmatrix} \sim N \left[ \begin{pmatrix} 0 \\ -\lambda \left( \frac{1}{N} \sum_{i=1}^N y_{i,1}(y_{i,1} - \mu_i) \right) \end{pmatrix}, \sigma^2 \left( \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 \right) \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right].$$

PROOF

See Appendix B.2.2.  $\xrightarrow{a}$  stands for “has large sample distribution”.

We observe that the Arellano-Bond GMM estimator for the conditional ARFE(1) model is also biased. Kruiniger (2000b) has developed a median unbiased minimum distance estimator for conditional AR(X)FE(1) models which is based on Arellano-Bond type moment conditions and which can accommodate exogenous regressors. Kiviet (1995) has proposed a biased corrected LSDV estimator for this model, while Alonso-Borrego and Arellano (1999) have argued that Symmetrically Normalised GMM estimators tend to have smaller bias than their conventional (GMM) counterparts. Their simulation results show that the optimal SN-GMM estimators for various versions of the model are less biased indeed, but in some cases the biases are still substantial. Moreover, it should be noted that their simulation study was based on data generated by a stationary AREC(1) model and the question arises whether different DGP’s would yield similar results.

Also in contrast to the moment conditions for the inclusive model, none of the moment conditions for the conditional model identify the unit root:

**Theorem 13** *The moment conditions corresponding to the conditional ARFE(1) model given in (14) and (15) do not identify the unit root  $r = 1$ .*

PROOF

Trivial, see Appendix B.2.3.

So the moment conditions for the conditional model contain no information when  $r$  is exactly equal to one.

It follows from theorem 13 that there exists no unit root test which is based on a GMM estimator for the conditional model that can also be shown to be consistent when  $|r| < 1$  without making additional assumptions. However, when  $r = 1$  there are three extra moment conditions available:  $E((y_{i,3} - ry_{i,2})\Delta y_{i,2}) = 0$ ,  $E(y_{i,1}(y_{i,2} - ry_{i,1})) = 0$  and  $E(y_{i,2} - ry_{i,1}) = 0$



<sup>16</sup>. The first moment condition is due to Arellano and Bover (1995) and remains valid under stationarity. GMM estimators that exploit (versions of) some of these moment conditions can in principle be used for testing the unit root hypothesis. When  $r = 1$  the distribution of some of these GMM estimators depend on the initial conditions. However, there also exist statistics that can be used for testing the unit root hypothesis whose distribution does not depend on the initial observations when  $r = 1$ , for instance the Arellano-Bover GMM estimator (see also section 4.1.3, theorem 10), or the corrected LSDV estimator (see Harris and Tzavalis (1999)), which in fact can also be viewed as a GMM estimator.

## 5 Conclusions

In this paper we have discussed GMM based inference procedures for two types of dynamic panel data models with fixed effects: the conditional ARFE(1) model and the inclusive ARFE(1) model. In particular, we have considered the distributional properties of various GMM estimators when the autoregressive parameter  $r$  is close to or equal to one, as well as unit root tests.

First we presented a two-step Optimal Linear GMM estimator for the inclusive model which is asymptotically equivalent to the non-linear Optimal GMM estimator due to Ahn and Schmidt (1997).

Then we examined for both versions of the ARFE(1) model which moment conditions are weak when the true value of the autoregressive parameter is close to the unit root. We found that all the first and second moment conditions for the conditional model are weak

---

<sup>16</sup>There is still one other moment condition available when  $r = 1$ , e.g.  $E((r - 1)(\Delta y_{i,2})^2 - 2\Delta y_{i,2}\Delta y_{i,3}) = 0$  (cf (11) and (16)). Given  $E((y_{i,3} - ry_{i,2})\Delta y_{i,2}) = 0$ , the moment condition  $E((r - 1)(\Delta y_{i,2})^2 - 2\Delta y_{i,2}\Delta y_{i,3}) = 0$  is equivalent to  $E((r - 1)(y_{i,2}^2 - y_{i,1}^2)) = 0$ . When  $|r| < 1$  and  $E(y_{i,2}^2 - y_{i,1}^2) \neq 0$  adding the latter moment condition to the other moment conditions will result in a GMM estimate of  $r$  that is closer to 1 when using a diagonal weighting matrix. Therefore a unit root test should not be based on a GMM estimator that exploits this moment condition unless  $E(y_{i,2}^2 - y_{i,1}^2) = 0$  under the alternative hypothesis (as is the case under stationarity, see also section 4.1).

in the traditional sense near the unit root. We have also derived an approximation to the finite sample distribution of the Arellano-Bond estimator for the conditional model using local-to-zero asymptotics in the spirit of Staiger and Stock (1997).

In the case of the inclusive model the Arellano-Bond moment conditions are no longer weak near the unit root. However, the finite sample distributions of the Arellano-Bond and Arellano-Bover estimators for the inclusive model are affected by fact that the variances (and the covariances) of the cross-products of the instrument(s) with the regressor(s) and the dependent variable respectively explode when  $r$  approaches unity. By choosing a parameter sequence such that these variances become  $O(1)$ , i.e.  $r = 1 - \lambda/N$ , we have obtained an approximation to the finite sample distributions of the Arellano-Bond and Arellano-Bover estimators for the inclusive model which conform well to the Monte Carlo evidence for these estimators. The approximation for the Arellano-Bover estimator suggests that this estimator is (median) unbiased but has a larger variance than what is implied by conventional fixed-parameter asymptotics.

The stationarity assumption is very helpful for the estimation of  $r$  when  $r$  is close to one. This assumption implies two extra moment conditions for  $r$  relative to the moment conditions that are also valid under the conditional model. Unlike the other moment conditions these two moment conditions remain informative close to the unit root.

We conclude that one should not use standard GMM estimators for the conditional model when the value of  $r$  is close to one but that one can still use various GMM estimators for the inclusive model in this case. Kruiniger (2000b) has developed a median unbiased minimum distance estimator for the conditional ARFE(1) model.

We have also considered the unit root case. First, we have found that the Optimal Linear GMM estimator is superefficient when  $r$  equals one. This is due to a discontinuity in the derivative of one of the moment conditions. The moment condition in question only identifies  $r$  when  $r$  is equal to one. On the other hand none of the moment conditions for the conditional ARFE(1) model identify  $r$  when  $r$  is equal to one. We have also suggested various GMM based tests of the unit root against both versions of the ARFE(1) model.

## A PROOFS OF THE RESULTS IN SECTION 2

Let us define

$$y_i^2 \equiv [y_{i,1}^2 \quad y_{i,1}y_{i,2} \quad y_{i,1}y_{i,3} \quad y_{i,2}^2 \quad y_{i,2}y_{i,3} \quad y_{i,3}^2]'$$

$$G = \begin{bmatrix} r & -1-r & 1 & 0 & 0 & 0 \\ 0 & r & -1 & -r & 1 & 0 \\ 0 & r & 0 & -1 & -r & 1 \\ r-1 & -2r & 2 & r+1 & -2 & 0 \end{bmatrix}$$

For  $T = 3$ , the Linear GMM estimator exploits  $E(Gy_i^2) = 0$ . The four elements of  $E(Gy_i^2) = 0$  are (5), (7), (8) for  $T = 3$ , and  $E((r-1)(\Delta y_{i,2})^2 - 2\Delta y_{i,2}\Delta y_{i,3}) = 0$  (11), respectively. It can easily be seen that  $\text{rank}(G) = 4$  unless  $r = 1$ .

We will first show that the non-linear moment condition proposed by Ahn and Schmidt (1997)

$$E(y_{i,1}^2 + y_{i,2}\Delta\varepsilon_{i,3}/(1-r^2) - (y_{i,3} - ry_{i,2})(y_{i,2} - ry_{i,1})/(1-r)^2) = 0 \quad (9)$$

is redundant relative to  $E(Gy_i^2) = 0$ .

Let us define  $H = [r^3 - r^2 - r + 1 \quad -r^3 - 2r^2 + r \quad r^2 + r \quad 2r^2 + r - 1 \quad -2r \quad 0]$ ,  $K = [G' \quad H']'$  and  $k(r)' = [r(1-r) \quad 2(1+r) \quad 0 \quad 1 \quad 1]$ . Observe that (9) is equivalent to  $E(Hy_i^2) = 0$ . Since  $K'k(r) = 0$ , we find that  $\text{rank}(K) = 4$ , so (9) is redundant relative to  $E(Gy_i^2) = 0$ .

If we replace (11) in  $E(Gy_i^2) = 0$  by (9), and if we use an optimal weighting matrix, we obtain an asymptotically equivalent estimator because the rank of the matrix that is obtained after replacing the last row of  $G$  by  $H$  is still 4.

### A.1 Proof of theorem 1:

First, we will show that (11) is not redundant relative to (5), (7), (8). Next, we will show that (8) is not redundant relative to (5), (7). Finally, we will show that (7) is not redundant relative to (5).

Note that  $T = 3$ , and  $-1 < r < 1$ .

Let  $m(r) = m = (m_1(r) \ m_2(r) \ m_3(r) \ m_4(r))'$  where

$$\begin{aligned} m_1(r) &= \varepsilon_{i,3} \Delta y_{i,2} = (y_{i,3} - r y_{i,2}) \Delta y_{i,2} \\ m_2(r) &= y_{i,1} \Delta \varepsilon_{i,3} = y_{i,1} (\Delta y_{i,3} - r \Delta y_{i,2}) \\ m_3(r) &= y_{i,3} \varepsilon_{i,3} - y_{i,2} \varepsilon_{i,2} = y_{i,3} (y_{i,3} - r y_{i,2}) - y_{i,2} (y_{i,2} - r y_{i,1}) \\ m_4(r) &= (r-1) (\Delta y_{i,2})^2 - 2 \Delta y_{i,2} \Delta y_{i,3} \end{aligned}$$

When  $T = 3$ , the Linear GMM estimator exploits  $E(m(r)) = 0$ .

$$\text{We have } E\left(\frac{dm}{dr}\right) = \begin{bmatrix} -\frac{1}{1+r} \sigma^2 \\ \frac{1}{1+r} \sigma^2 \\ 0 \\ \frac{2}{1+r} \sigma^2 \end{bmatrix}.$$

Under normality and stationarity  $E((\text{vec}(Z_i - \Omega))(\text{vec}(Z_i - \Omega))') = E((\text{vec} Z_i)(\text{vec} Z_i)') - (\text{vec} \Omega)(\text{vec} \Omega)'$   $= (\Omega \otimes \Omega)(I_{(T-1)^2} + P_{T-1})$ , where  $Z_i = D y_i y_i' D'$ ,  $\Omega = \Omega(r, \sigma^2) = E(Z_i) =$

$$\sigma^2 D V(r) D', P \text{ is the commutation matrix, and } I_{(T-1)^2} + P_{T-1} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

It follows that

$$E((\text{vech}(Z_i - \Omega))(\text{vech}(Z_i - \Omega))') = \begin{bmatrix} \frac{8}{(1+r)^2} \sigma^2 & -4 \frac{1-r}{(1+r)^2} \sigma^2 & 2 \frac{(1-r)^2}{(1+r)^2} \sigma^2 \\ -4 \frac{1-r}{(1+r)^2} \sigma^2 & \frac{5-2r+r^2}{(1+r)^2} \sigma^2 & -4 \frac{1-r}{(1+r)^2} \sigma^2 \\ 2 \frac{(1-r)^2}{(1+r)^2} \sigma^2 & -4 \frac{1-r}{(1+r)^2} \sigma^2 & \frac{8}{(1+r)^2} \sigma^2 \end{bmatrix}.$$

Using this result, we obtain

$$E(mm') = \begin{bmatrix} \frac{2}{1+r} & -\frac{1}{1+r} & \frac{r}{1+r} & -\frac{4}{1+r} \\ -\frac{1}{1+r} & 2b + \frac{2}{1-r^2} & 2b + \frac{r}{1-r} & \frac{2(1-r)}{1+r} \\ \frac{r}{1+r} & 2b + \frac{r}{1-r} & 4 + 2b + \frac{2r^2}{1-r^2} & -2 \\ -\frac{4}{1+r} & \frac{2(1-r)}{1+r} & -2 & \frac{4(11-10r+3r^2)}{(1+r)^2} \end{bmatrix} \sigma^4,$$

where  $b = \frac{1}{N} \sum_{i=1}^N \frac{\mu_i^2}{\sigma^2} \geq 0$ .

Let  $\hat{r}_{GMM\#}$  denote the Optimal GMM estimator that is based on the first  $\#$  moment conditions comprised in  $E(m) = 0$ . Note that  $\hat{r}_{LGMM} = \hat{r}_{GMM4}$ . Then

$$\text{Var}(\hat{r}_{GMM3}) = -\frac{1}{2(br^2+2br-2-3b)} (1+r) (br^3 - 7br^2 + 2r - 5br + 6 + 11b),$$

$$\text{Var}(\hat{r}_{LGMM}) = -\frac{1}{2(5br^3-5br^2-r^2-53br-13r+35+53b-r^3)} (1+r) \times (6br^4 - r^4 + 4r^3 - 56br^3 + 18r^2 + 92br^2 + 152br + 4r - 105 - 194b), \text{ and}$$

$$\text{Var}(\hat{r}_{LGMM}) - \text{Var}(\hat{r}_{GMM3}) = -\frac{1}{2} (1+r)^2 \frac{(r-1)(br^2-2br-2r+b)^2}{(5br^3-5br^2-r^2-53br-13r+35+53b-r^3)(br^2+2br-2-3b)}.$$

We will first show that  $(5br^3 - 5br^2 - r^2 - 53br - 13r + 35 + 53b - r^3) > 0$  :

$$5br^3 - 5br^2 - 53br + 53b = (r - 1)(5r^2 - 53)b \geq 0, \text{ and}$$

$$-r^2 - 13r + 35 - r^3 > 0.$$

Next we will show that  $(br^2 + 2br - 2 - 3b) < 0$  :

$$br^2 + 2br - 3b = (r + 3)(r - 1)b \leq 0, \text{ and } -2 < 0.$$

Finally, we will examine whether  $br^2 - 2br - 2r + b$  can equal 0:

When  $b > 0$ , we can write  $br^2 - 2br - 2r + b = b(r^2 - 2(\frac{b+1}{b})r + 1)$ . Since  $4(\frac{b+1}{b})^2 - 4 > 0$ ,  $br^2 - 2br - 2r + b = 0$  will have two real solutions. Both solutions will be positive, with one solution being smaller than 1 and the other solution being larger than 1.

When  $b = 0$ ,  $br^2 - 2br - 2r + b = -2r$ . Now,  $-2r = 0$  iff  $r = 0$ .

We conclude that  $Var(\widehat{r}_{GMM3}) > Var(\widehat{r}_{LGMM})$  unless  $br^2 - 2br - 2r + b = 0$ .

Since  $Var(\widehat{r}_{GMM2}) = \frac{1}{2(-b+br^2-1)}(1+r)(4br^2 - r - 3 - 4b)$ ,

$Var(\widehat{r}_{GMM3}) = -\frac{1}{2(br^2+2br-2-3b)} \times (1+r)(br^3 - 7br^2 + 2r - 5br + 6 + 11b)$ , we find that

$$Var(\widehat{r}_{GMM3}) - Var(\widehat{r}_{GMM2}) = -\frac{1}{2}(1+r)^2 b^2 \frac{(r-1)^4}{(br^2+2br-2-3b)(-b+br^2-1)}.$$

First we will show that  $(br^2 + 2br - 2 - 3b) < 0$  :

$$br^2 + 2br - 3b = (r + 3)(r - 1)b \leq 0 \text{ and } -2 < 0.$$

Next we will show that  $(-b + br^2 - 1) < 0$  :

$$-b + br^2 = b(r - 1)(1 + r) \leq 0 \text{ and } -1 < 0.$$

We conclude that  $Var(\widehat{r}_{GMM2}) > Var(\widehat{r}_{GMM3})$  unless  $b = 0$  ( $\mu_i^2 = 0 \forall i \in \{1, \dots, N\}$ ).

Notice that the third moment condition,  $E(m_3(r)) = 0$ , does not identify  $r$  when  $r < 1$ . Nonetheless including  $E(m_3(r)) = 0$  in the set of moment conditions generally leads to more efficient estimators.

Let  $\widehat{r}_{GMMAB}$  be the GMM estimator that is based on the Arrelano-Bond moment condition  $E(m_2(r)) = E(y_{i,1}\Delta\varepsilon_{i,3}) = 0$ . Then  $Var(\widehat{r}_{GMMAB}) = (1+r)^2(2b + \frac{2}{1-r^2})$ ,

$Var(\widehat{r}_{GMM2}) = \frac{1}{2(-b+br^2-1)}(1+r)(4br^2 - r - 3 - 4b)$ , and

$$Var(\widehat{r}_{GMM2}) - Var(\widehat{r}_{GMMAB}) = -\frac{1}{2}(1+r)^3 \frac{(2rb-2b-1)^2}{(-b+br^2-1)(r-1)}.$$

We will first show that  $(-b + br^2 - 1) < 0$  :

$$-b + br^2 = (r - 1)(1 + r)b \leq 0 \text{ and } -1 < 0.$$

Next we will show that  $2rb - 2b - 1 < 0$  :

$$2rb - 2b - 1 = 2(r - 1)b \leq 0 \text{ and } -1 < 0.$$

We conclude that  $Var(\widehat{r}_{GMMAB}) > Var(\widehat{r}_{GMM2})$ .  $\square$

## B PROOF OF THE RESULTS IN SECTION 4

### B.1 Proofs of results in section 4.1

#### B.1.1 Various results:

$m(r) = m = (m_1(r) \ m_2(r) \ m_3(r) \ m_4(r))'$ , where  $m_1(r)$ ,  $m_2(r)$ ,  $m_3(r)$ , and  $m_4(r)$  have already been defined in section 4.1 (and also in appendix A).

Properties of  $E(m(r)m(r)')$ :

$$\text{When } |r| < 1, E(m(r)m(r)') = \begin{bmatrix} \frac{2}{1+r} & -\frac{1}{1+r} & \frac{r}{1+r} & -\frac{4}{1+r} \\ -\frac{1}{1+r} & 2b + \frac{2}{1-r^2} & 2b + \frac{r}{1-r} & \frac{2(1-r)}{1+r} \\ \frac{r}{1+r} & 2b + \frac{r}{1-r} & 4 + 2b + \frac{2r^2}{1-r^2} & -2 \\ -\frac{4}{1+r} & \frac{2(1-r)}{1+r} & -2 & \frac{4(11-10r+3r^2)}{(1+r)^2} \end{bmatrix} \sigma^4.$$

$$E(m(1)m(1)' | y_{1,1}, \dots, y_{N,1}) = \begin{bmatrix} \sigma^4 & 0 & \sigma^4 & -2\sigma^4 \\ 0 & 2\sigma^2 \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 & 2\sigma^2 \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 & 0 \\ \sigma^4 & 2\sigma^2 \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 & 5\sigma^4 + 2\sigma^2 \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 & -2\sigma^4 \\ -2\sigma^4 & 0 & -2\sigma^4 & 4\sigma^4 \end{bmatrix}.$$

Properties of  $E(\frac{dm(r)}{dr})$ :

$$\lim_{r \uparrow 1} E(\frac{dm(r)}{dr}) = \lim_{r \uparrow 1} \begin{bmatrix} -\frac{1}{1+r}\sigma^2 \\ \frac{1}{1+r}\sigma^2 \\ 0 \\ \frac{2}{1+r}\sigma^2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\sigma^2 \\ \frac{1}{2}\sigma^2 \\ 0 \\ \sigma^2 \end{bmatrix}, \quad E(\frac{dm(r)}{dr} |_{r=1}) = \begin{bmatrix} -\sigma^2 \\ 0 \\ -\sigma^2 \\ \sigma^2 \end{bmatrix}.$$

Remember that  $\tilde{m}(r) = (m_2(r) \ m_3(r) \ m_4(r) \ m_5(r))'$ .

Properties of  $E(\tilde{m}(r)\tilde{m}(r)')$  and  $[E(\tilde{m}(r)\tilde{m}(r)')]^{-1}$ :

$$\text{Let } V(r, b) = E(\tilde{m}(r)\tilde{m}(r)') = \begin{bmatrix} 2\frac{-b+br^2-1}{-1+r^2} & \frac{2br-2b-r}{r-1} & -\frac{2r-1}{1+r} & -\frac{2r}{1+r} \\ \frac{2br-2b-r}{r-1} & 2\frac{-2+r^2-b+br^2}{-1+r^2} & -2 & -\frac{2}{1+r} \\ -\frac{2r-1}{1+r} & -2 & 4\frac{11-10r+3r^2}{(1+r)^2} & 12\frac{3-4r+r^2}{(1+r)^2} \\ -\frac{2r}{1+r} & -\frac{2}{1+r} & 12\frac{3-4r+r^2}{(1+r)^2} & 12\frac{3-4r+r^2}{(1+r)^2} \end{bmatrix} \sigma^4,$$

and  $W(r, b) = V^{-1}(r, b)$ .

We find that  $\lim_{r \uparrow 1} W(r, b) = \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & -\frac{1}{8} & \frac{9}{40} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{8} & -\frac{1}{40} \\ -\frac{1}{8} & \frac{1}{8} & \frac{5}{16} & -\frac{1}{5} \\ \frac{9}{40} & -\frac{1}{40} & -\frac{5}{16} & \infty \end{bmatrix} \sigma^{-4}$ .

For  $r$  local to unity we have  $W(r, b) \approx \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & -\frac{1}{8} & \frac{9}{40} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{8} & -\frac{1}{40} \\ -\frac{1}{8} & \frac{1}{8} & \frac{5}{16} & -\frac{1}{5} \\ \frac{9}{40} & -\frac{1}{40} & -\frac{5}{16} & \frac{1}{5(1-r)} \end{bmatrix} \sigma^{-4}$ .

Define  $\widetilde{W}(r, b) = (W_{11}(r, b) + W_{22}(r, b) + W_{33}(r, b) + W_{44}(r, b))^{-1} W(r, b)$ .

Then  $\lim_{r \uparrow 1} \widetilde{W}(r, b) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ .

Now  $E(\widetilde{m}(1)\widetilde{m}(1)' | y_{1,1}, \dots, y_{N,1}) = \begin{bmatrix} 2\sigma^2 \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 & 2\sigma^2 \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 & 0 & 0 \\ 2\sigma^2 \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 & 5\sigma^4 + 2\sigma^2 \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 & -2\sigma^4 & 0 \\ 0 & -2\sigma^4 & 4\sigma^4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ ,

so when  $r = 1$  the normalised optimal weighting matrix  $\widetilde{W}(1, b) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ .

We conclude that  $\lim_{r \uparrow 1} \widetilde{W}(r, b) = \widetilde{W}(1, b)$ .

### B.1.2 Proof of theorem 5 (local-to-unity asymptotics, $T = 3$ ):

$y_{i,t} = ry_{i,t-1} + (1-r)\mu_i + \varepsilon_{i,t}$  and we assume stationarity. We will derive the local-to-unity asymptotic distributions of the following estimators:

$$\widehat{r}_{ARBOV} = \frac{\frac{1}{N} \sum_{i=1}^N y_{i,3} \Delta y_{i,2}}{\frac{1}{N} \sum_{i=1}^N y_{i,2} \Delta y_{i,2}} = r + \frac{\frac{1}{N} \sum_{i=1}^N (\varepsilon_{i,3} + \mu_i) \Delta y_{i,2}}{\frac{1}{N} \sum_{i=1}^N y_{i,2} \Delta y_{i,2}}$$

$$\widehat{r}_{ARBOND} = \frac{\frac{1}{N} \sum_{i=1}^N y_{i,1} \Delta y_{i,3}}{\frac{1}{N} \sum_{i=1}^N y_{i,1} \Delta y_{i,2}} = r + \frac{\frac{1}{N} \sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3}}{\frac{1}{N} \sum_{i=1}^N y_{i,1} \Delta y_{i,2}}$$

Using  $u_t = \sum_{s=0}^{\infty} r^s \varepsilon_{t-s}$ , we obtain the following results:

$$E(u_{i,1}^2) = \frac{\sigma^2}{1-r^2}, \quad E(u_{i,1} \Delta u_{i,2}) = -\frac{\sigma^2}{1+r}, \quad E[(\Delta u_{i,2})^2] = \frac{2\sigma^2}{1+r},$$

$$E(u_{i,1}^4) = E\left(\sum_{s=0}^{\infty} r^{4s} \varepsilon_{i,1-s}^4 + 6 \sum_{s=0}^{\infty} \sum_{t=s+1}^{\infty} r^{2(s+t)} \varepsilon_{i,1-s}^2 \varepsilon_{i,1-t}^2\right) = \frac{3\sigma^4}{1-r^4} + 6\sigma^4 r^2 \sum_{s=0}^{\infty} \frac{r^{4s}}{1-r^2} = \frac{3\sigma^4}{1-r^4} + \frac{6\sigma^4 r^2}{(1-r^2)(1-r^4)} = \frac{3\sigma^4}{(1-r^2)^2},$$

$$E[(\Delta u_{i,2})^4] = E[(r-1)^4 u_{i,1}^4 + 6(r-1)^2 u_{i,1}^2 \varepsilon_{i,2}^2 + \varepsilon_{i,2}^4] = \frac{3(1-r)^2 \sigma^4}{(1+r)^2} + \frac{6(1-r)\sigma^4}{1+r} + 3\sigma^4 = \frac{3(1-r)^2 + 6(1-r)(1+r) + 3(1+r)^2}{(1+r)^2} \sigma^4 = \frac{12\sigma^4}{(1+r)^2},$$

$$E[u_{i,1} (\Delta u_{i,2})^3] = E[(r-1)^3 u_{i,1}^4 + 3(r-1) u_{i,1}^2 \varepsilon_{i,2}^2] = -\frac{3(1-r)\sigma^4}{(1+r)^2} - \frac{3\sigma^4}{1+r} = -\frac{6\sigma^4}{(1+r)^2},$$

$$E[u_{i,1}^2 (\Delta u_{i,2})^2] = E[(r-1)^2 u_{i,1}^4 + u_{i,1}^2 \varepsilon_{i,2}^2] = \frac{3\sigma^4}{(1+r)^2} + \frac{\sigma^4}{1-r^2} = \frac{2(2-r)\sigma^4}{(1+r)^2(1-r)},$$

$$E(y_{i,1} \Delta \varepsilon_{i,3}) = 0,$$

$$E[(\varepsilon_{i,3} + (1-r)\mu_i) \Delta y_{i,2}] = E[(\varepsilon_{i,3} + (1-r)\mu_i) \Delta u_{i,2}] = 0,$$

$$E(y_{i,1} \Delta y_{i,2}) = E[(u_{i,1} + \mu_i) \Delta u_{i,2}] = -\frac{\sigma^2}{1+r},$$

$$E(y_{i,2} \Delta y_{i,2}) = E(\Delta y_{i,2} \Delta y_{i,2} + y_{i,1} \Delta y_{i,2}) = E(\Delta u_{i,2} \Delta u_{i,2} + y_{i,1} \Delta y_{i,2}) = \frac{\sigma^2}{1+r},$$

$$\text{Var}(y_{i,1} \Delta \varepsilon_{i,3}) = E[(y_{i,1} \Delta \varepsilon_{i,3})^2] = E\{[(u_{i,1} + \mu_i) \Delta \varepsilon_{i,3}]^2\} = \frac{2\sigma^4}{1-r^2} + 2\mu_i^2 \sigma^2 = \frac{2\sigma^2[\sigma^2 + \mu_i^2(1-r^2)]}{(1+r)(1-r)},$$

$$\text{Var}[(\varepsilon_{i,3} + (1-r)\mu_i) \Delta y_{i,2}] = E\{[(\varepsilon_{i,3} + (1-r)\mu_i) \Delta u_{i,2}]^2\} = \frac{2\sigma^2[\sigma^2 + \mu_i^2(1-r^2)]}{1+r},$$

$$\text{Var}(y_{i,1} \Delta y_{i,2}) = E[(u_{i,1} \Delta u_{i,2} + \mu_i \Delta u_{i,2} + \frac{\sigma^2}{1+r})^2] = E[(u_{i,1} \Delta u_{i,2})^2 - \frac{2\sigma^4}{(1+r)^2} + \frac{2\sigma^2 \mu_i^2}{1+r} + \frac{\sigma^4}{(1+r)^2}] = \frac{2(2-r)\sigma^4}{(1+r)^2(1-r)} + \frac{\sigma^2(2\mu_i^2(1+r) - \sigma^2)}{(1+r)^2} = \frac{\sigma^2(2\mu_i^2(1-r^2) + (3-r)\sigma^2)}{(1+r)^2(1-r)},$$

$$\text{Var}(y_{i,2} \Delta y_{i,2}) = E[(u_{i,1} \Delta u_{i,2} + (\Delta u_{i,2})^2 + \mu_i \Delta u_{i,2} - \frac{\sigma^2}{1+r})^2] =$$

$$E[(u_{i,1} \Delta u_{i,2})^2 + (\Delta u_{i,2})^4 + (\mu_i \Delta u_{i,2})^2 + \frac{\sigma^4}{(1+r)^2} + 2u_{i,1} (\Delta u_{i,2})^3 - \frac{2\sigma^2}{1+r} (u_{i,1} \Delta u_{i,2}) - \frac{2\sigma^2}{1+r} (\Delta u_{i,2})^2] = \frac{2(2-r)\sigma^4}{(1+r)^2(1-r)} + \frac{12\sigma^4}{(1+r)^2} + \frac{2\sigma^2 \mu_i^2}{1+r} + \frac{\sigma^4}{(1+r)^2} - \frac{12\sigma^4}{(1+r)^2} + \frac{2\sigma^4}{(1+r)^2} - \frac{4\sigma^4}{(1+r)^2} = \frac{2(2-r)\sigma^4}{(1+r)^2(1-r)} + \frac{\sigma^2[2\mu_i^2(1+r) - 1]}{(1+r)^2} = \frac{\sigma^2[(3-r)\sigma^2 + 2\mu_i^2(1-r^2)]}{(1+r)^2(1-r)},$$

$$\text{Cov}(y_{i,1} \Delta \varepsilon_{i,3}, y_{i,1} \Delta y_{i,2}) = \text{Cov}(y_{i,1} \Delta \varepsilon_{i,3}, y_{i,1} \Delta u_{i,2}) = \text{Cov}(y_{i,1} \Delta \varepsilon_{i,3}, u_{i,1} \Delta u_{i,2}) +$$

$$\text{Cov}(y_{i,1} \Delta \varepsilon_{i,3}, \mu_i \Delta u_{i,2}) = \text{Cov}(u_{i,1} \Delta \varepsilon_{i,3}, u_{i,1} \Delta u_{i,2}) + \text{Cov}(\mu_i \Delta \varepsilon_{i,3}, \mu_i \Delta u_{i,2}) =$$



$E(u_{i,1}^2 \Delta \varepsilon_{i,3} \Delta u_{i,2}) + \mu_i^2 E(\Delta \varepsilon_{i,3} \Delta u_{i,2}) = -E(u_{i,1}^2 \varepsilon_{i,2}^2) - \mu_i^2 E(\varepsilon_{i,2}^2) = -\frac{\sigma^2[\sigma^2 + \mu_i^2(1-r^2)]}{(1-r)(1+r)}$ , and

$$\text{Cov}[(\varepsilon_{i,3} + (1-r)\mu_i)\Delta y_{i,2}, y_{i,2}\Delta y_{i,2}] = \text{Cov}(\varepsilon_{i,3}\Delta u_{i,2}, y_{i,2}\Delta u_{i,2}) +$$

$$\text{Cov}[(1-r)\mu_i\Delta u_{i,2}, y_{i,2}\Delta u_{i,2}] = E[(1-r)(\mu_i\Delta u_{i,2})^2] = \frac{2\sigma^2\mu_i^2(1-r)}{1+r}.$$

Let us define  $X_1 = \frac{1}{N} \sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3}$ ,

$$X_2 = \frac{1}{N} \sum_{i=1}^N y_{i,1} \Delta y_{i,2},$$

$$X_3 = N^{-0.5} \sum_{i=1}^N [(\varepsilon_{i,3} + (1-r)\mu_i)\Delta y_{i,2}], \text{ and}$$

$$X_4 = \frac{1}{N} \sum_{i=1}^N y_{i,2} \Delta y_{i,2}.$$

Let us consider the sequence  $r = 1 - \frac{\lambda}{N}$  while keeping the  $\mu'_i$ 's fixed.

$$\text{Then } \lim_{N \rightarrow \infty} \text{Var}(X_1) = \frac{\sigma^4}{\lambda}; X_1 \xrightarrow{d} N(0, \frac{\sigma^4}{\lambda}),$$

$$\lim_{N \rightarrow \infty} \text{Var}(X_2) = \frac{\sigma^4}{2\lambda}; X_2 \xrightarrow{d} N(-\frac{\sigma^2}{2}, \frac{\sigma^4}{2\lambda}),$$

$$\lim_{N \rightarrow \infty} \text{Cov}(X_1, X_2) = -\frac{\sigma^4}{2\lambda},$$

$$\text{plim}_{N \rightarrow \infty} N^{-0.5} X_3 = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N [(\varepsilon_{i,3} + (1-r)\mu_i)\Delta y_{i,2}] = 0,$$

$$\lim_{N \rightarrow \infty} \text{Var}(X_3) = \sigma^4; X_3 \xrightarrow{d} N(0, \sigma^4),$$

$$\lim_{N \rightarrow \infty} \text{Var}(X_4) = \frac{\sigma^4}{2\lambda}; X_4 \xrightarrow{d} N(\frac{\sigma^2}{2}, \frac{\sigma^4}{2\lambda}), \text{ and}$$

$$\lim_{N \rightarrow \infty} \text{Cov}(X_3, X_4) = 0.$$

It follows that  $\text{plim}_{N \rightarrow \infty} \hat{r}_{ARBOV} = r$ ,

$$\hat{r}_{ARBOND} \xrightarrow{d} r + \frac{\tilde{X}_1}{\tilde{X}_2}, \text{ where}$$

$$\begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix} \sim N \left[ -\frac{\sigma^2}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \frac{\sigma^4}{2\lambda} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right], \text{ and}$$

$$\sqrt{N}(\hat{r}_{ARBOV} - r) \xrightarrow{d} \frac{\tilde{X}_3}{\tilde{X}_4}, \text{ where}$$

$$\begin{bmatrix} \tilde{X}_3 \\ \tilde{X}_4 \end{bmatrix} \sim N \left[ \frac{\sigma^2}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \sigma^4 & 0 \\ 0 & \frac{\sigma^4}{2\lambda} \end{pmatrix} \right]. \square$$

Alternatively, let us consider the model  $y_{i,t} = ry_{i,t-1} + \eta_i + \varepsilon_{i,t}$  with stationarity, and the sequence  $r = 1 - (\frac{\lambda}{N})^{0.5}$ , and let us now keep the  $\eta'_i$ 's fixed. For convenience we assume that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \eta_i^2 = \eta^2.$$

Define  $Y_3 = N^{-0.5} \sum_{i=1}^N [(\varepsilon_{i,3} + \eta_i)\Delta y_{i,2}]$ , and  $Y_i = X_i$ ,  $i = 1, 2, 4$ . Then we have

$$\lim_{N \rightarrow \infty} \text{Var}(Y_1) = \frac{2\sigma^2\eta^2}{\lambda}; Y_1 \xrightarrow{d} N(0, \frac{2\sigma^2\eta^2}{\lambda}),$$

$$\lim_{N \rightarrow \infty} \text{Var}(Y_2) = \frac{\sigma^2\eta^2}{\lambda}; Y_2 \xrightarrow{d} N(-\frac{\sigma^2}{2}, \frac{\sigma^2\eta^2}{\lambda}),$$

$$\lim_{N \rightarrow \infty} \text{Cov}(Y_1, Y_2) = -\frac{\sigma^2\eta^2}{\lambda},$$

$$\begin{aligned} \text{plim}_{N \rightarrow \infty} N^{-0.5} Y_3 &= \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N [(\varepsilon_{i,3} + \eta_i) \Delta y_{i,2}] = 0, \\ \lim_{N \rightarrow \infty} \text{Var}(Y_3) &= \sigma^2(\sigma^2 + \eta^2); Y_3 \xrightarrow{d} N(0, \sigma^2[\sigma^2 + \eta^2]), \\ \lim_{N \rightarrow \infty} \text{Var}(Y_4) &= \frac{\sigma^2 \eta^2}{\lambda}; Y_4 \xrightarrow{d} N\left(\frac{\sigma^2}{2}, \frac{\sigma^2 \eta^2}{\lambda}\right), \text{ and} \\ \lim_{N \rightarrow \infty} \text{Cov}(Y_3, Y_4) &= \lim_{N \rightarrow \infty} \sigma^2 \eta^2 \lambda^{-0.5}. \end{aligned}$$

It follows that  $\text{plim}_{N \rightarrow \infty} \widehat{r}_{ARBOV} = r$ ,

$$\widehat{r}_{ARBOND} \xrightarrow{d} r + \frac{\tilde{Y}_1}{\tilde{Y}_2}, \quad \text{where}$$

$$\begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \end{bmatrix} \sim N \left[ -\frac{\sigma^2}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \frac{\sigma^2 \eta^2}{\lambda} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right], \text{ and}$$

$$\sqrt{N}(\widehat{r}_{ARBOV} - r) \xrightarrow{d} \frac{\tilde{Y}_3}{\tilde{Y}_4}, \quad \text{where}$$

$$\begin{bmatrix} \tilde{Y}_3 \\ \tilde{Y}_4 \end{bmatrix} \sim N \left[ \frac{\sigma^2}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \sigma^2[\sigma^2 + \eta^2] & \sigma^2 \eta^2 \lambda^{-0.5} \\ \sigma^2 \eta^2 \lambda^{-0.5} & \frac{\sigma^2 \eta^2}{\lambda} \end{pmatrix} \right].$$

### B.1.3 Proof of theorem 6 (local-to-unity asymptotics, $T > 3$ ):

We will derive the local-to-unity asymptotic distribution of the optimal Arellano-Bond GMM estimator for arbitrary  $T > 3$  and show that locally-to-unity this estimator is asymptotically biased downwards,  $\lim_{N \rightarrow \infty} E(\widehat{r}_{ARBOND}) < r$ , if we choose the parameter sequence  $r = 1 - \frac{\lambda}{N}$ .

It follows from arguments in Arellano-Bover (1995) that the optimal Arellano-Bond GMM estimator is equal to an optimal GMM estimator where the fixed effects are eliminated from the model by applying the Helmert transformation, that is by taking forward orthogonal deviations, instead of first-differences. The Helmert transformation of  $y_{i,t}$  is  $\tilde{y}_{i,t} = \left(\frac{T-t}{T-t+1}\right)^{0.5} \times \left[ y_{i,t} - \frac{1}{T-t} \sum_{s=t+1}^T y_{i,s} \right]$ .

The advantage of applying the Helmert transformation matrix,  $(DD')^{-1/2}D$ , instead of the first difference matrix,  $D$ , to the system of  $T - 1$  equations is that it preserves the orthogonality among the errors, as  $\sigma^2(DD')^{-1/2}DID'(DD')^{-1/2} = \sigma^2I$ , which will prove to be helpful.

Applying the Helmert transformation to the  $T - 1$  equations yields  $\tilde{y}_{i,t} = r\tilde{y}_{i,t-1} + \tilde{\varepsilon}_{i,t}$ ,  $t = 2, \dots, T - 1$ .

All lagged values of  $y_{i,t}$  are valid instruments for the  $t$ -th equation. If we let  $z_{i,t} = [y_{i,1} \dots y_{i,t-1}]'$ , we have the following set of moment conditions:  $E(z_{i,t}\tilde{\varepsilon}_{i,t}) = 0$ ,  $t = 2, \dots, T - 1$ .

The optimal weighting matrix is proportional to a block-diagonal matrix with typical diagonal block equal to  $[E(z_{i,t}z'_{i,t})]^{-1}$ . The optimal GMM estimator is therefore

$$\hat{r} = \frac{\sum_{t=2}^{T-1} \tilde{y}'_{t-1} P_t \tilde{y}_t}{\sum_{t=2}^{T-1} \tilde{y}'_{t-1} P_t \tilde{y}_{t-1}} = r + \frac{\sum_{t=2}^{T-1} \tilde{y}'_{t-1} P_t \tilde{\varepsilon}_t}{\sum_{t=2}^{T-1} \tilde{y}'_{t-1} P_t \tilde{y}_{t-1}}, \quad \text{where } \tilde{y}_t = [y_{1,t} \dots y_{N,t}]', \tilde{\varepsilon}_t = [\varepsilon_{1,t} \dots \varepsilon_{N,t}]',$$

$$Z_t = [z_{1,t} \dots z_{N,t}]', \text{ and } P_t = Z_t(Z_t'Z_t)^{-1}Z_t'.$$

Now  $\text{plim}_{N \rightarrow \infty} N^2(Z_t'Z_t)^{-1} = \lim_{N \rightarrow \infty} N[E(z_{i,t}z_{i,t}')]^{-1}$ ,  $t = 2, \dots, T-1$ , if we consider the parameter sequence  $r = 1 - \frac{\lambda}{N}$ . (All limits are taken using this parameter sequence).

We note that  $\text{plim}_{N \rightarrow \infty} N^2(Z_t'Z_t)^{-1}$  is a PDS matrix, that is, *not* a singular matrix. For instance, when  $t = 3$ ,  $\text{plim}_{N \rightarrow \infty} \det[N^{-2}(Z_t'Z_t)] = \lim_{N \rightarrow \infty} \frac{1}{N} \frac{\sigma^4}{(1-r^2)} = \frac{\sigma^4}{2\lambda}$ .

Using similar arguments as in the proof of theorem 5,  $\frac{1}{N}\tilde{y}'_{t-1}Z_t$  and  $\frac{1}{N}Z_t'\tilde{\varepsilon}_t$  can be shown to converge in distribution to normally distributed random vectors if we consider the parameter sequence  $r = 1 - \frac{\lambda}{N}$ . It follows that  $\sum_{t=2}^{T-1} \tilde{y}'_{t-1}Z_t(Z_t'Z_t)^{-1}Z_t'\tilde{y}_{t-1}$  converges in distribution to a positive random variable.

We will now show that  $\lim_{N \rightarrow \infty} E(\tilde{y}'_{t-1}Z_t(Z_t'Z_t)^{-1}Z_t'\tilde{\varepsilon}_t) < 0$ .

We can write  $\lim_{N \rightarrow \infty} E(\tilde{y}'_{t-1}Z_t(Z_t'Z_t)^{-1}Z_t'\tilde{\varepsilon}_t) = \lim_{N \rightarrow \infty} E(\sum_{i=1}^N \tilde{y}_{i,t-1}z'_{i,t}(Z_t'Z_t)^{-1}z_{i,t}\tilde{\varepsilon}_{i,t}) = \lim_{N \rightarrow \infty} E(\sum_{i=1}^N z'_{i,t}(Z_t'Z_t)^{-1}z_{i,t}) \times \lim_{N \rightarrow \infty} E(\tilde{y}_{i,t-1}\tilde{\varepsilon}_{i,t})$ .

Now  $\lim_{N \rightarrow \infty} E(\sum_{i=1}^N z'_{i,t}(Z_t'Z_t)^{-1}z_{i,t}) = E[\text{tr}(I_{t-1})] = t-1 > 0$ .

$$\begin{aligned} \text{We also have } E(\tilde{y}_{i,t-1}\tilde{\varepsilon}_{i,t}) &= \left(\frac{T-t}{T-t+1}\right) E\left[(y_{i,t-1} - \frac{1}{T-t} \sum_{s=t}^{T-1} y_{i,s})(\varepsilon_{i,t} - \frac{1}{T-t} \sum_{s=t+1}^T \varepsilon_{i,s})\right] = \\ &= \left(\frac{1}{T-t+1}\right) E\left[-\sum_{k=0}^{T-1-t} r^k \varepsilon_{i,t}^2 + \frac{1}{T-t} \sum_{s=t+1}^{T-1} \sum_{k=0}^{T-1-s} r^k \varepsilon_{i,s}^2\right] = \\ &= \sigma^2 \left(\frac{1}{T-t+1}\right) \left(\frac{1}{T-t}\right) \left[-\sum_{s=t+1}^T \sum_{k=T-1-s+1}^{T-1-t} r^k\right]. \end{aligned}$$

It follows that  $\lim_{N \rightarrow \infty} E(\tilde{y}_{i,t-1}\tilde{\varepsilon}_{i,t}) = \lim_{r \uparrow 1} E(\tilde{y}_{i,t-1}\tilde{\varepsilon}_{i,t}) = -\frac{1}{2}\sigma^2$ .

We conclude that  $\lim_{N \rightarrow \infty} E(\sum_{t=2}^{T-1} \tilde{y}'_{t-1}Z_t(Z_t'Z_t)^{-1}Z_t'\tilde{\varepsilon}_t) < 0$ .

To complete the proof we write the numerator of  $\hat{r} - r$  as the sum of two terms:

$$\begin{aligned} \sum_{t=2}^{T-1} \tilde{y}'_{t-1}Z_t(Z_t'Z_t)^{-1}Z_t'\tilde{\varepsilon}_t &= \sum_{t=2}^{T-1} (\sum_{i=1}^N \tilde{y}_{i,t-1}z'_{i,t})(Z_t'Z_t)^{-1}(\sum_{i=1}^N z_{i,t}\tilde{\varepsilon}_{i,t}) = \\ &= \sum_{t=2}^{T-1} [\sum_{i=1}^N z'_{i,t}(Z_t'Z_t)^{-1}z_{i,t}(\frac{1}{T-t+1})(-\sum_{k=0}^{T-1-t} r^k \varepsilon_{i,t}^2 + \frac{1}{T-t} \sum_{s=t+1}^{T-1} \sum_{k=0}^{T-1-s} r^k \varepsilon_{i,s}^2)] + \\ &= \sum_{t=2}^{T-1} [\tilde{y}'_{t-1}Z_t(Z_t'Z_t)^{-1}Z_t'\tilde{\varepsilon}_t - \\ &= \sum_{i=1}^N z'_{i,t}(Z_t'Z_t)^{-1}z_{i,t}(\frac{1}{T-t+1})(-\sum_{k=0}^{T-1-t} r^k \varepsilon_{i,t}^2 + \frac{1}{T-t} \sum_{s=t+1}^{T-1} \sum_{k=0}^{T-1-s} r^k \varepsilon_{i,s}^2)]. \end{aligned}$$

The first term converges in probability to  $\lim_{N \rightarrow \infty} \sum_{t=2}^{T-1} E(\tilde{y}'_{t-1}Z_t(Z_t'Z_t)^{-1}Z_t'\tilde{\varepsilon}_t)$ . Since the numerator of  $\hat{r} - r$ ,  $\sum_{t=2}^{T-1} \tilde{y}'_{t-1}Z_t(Z_t'Z_t)^{-1}Z_t'\tilde{\varepsilon}_t$ , converges in distribution to a random variable with mean  $\lim_{N \rightarrow \infty} E(\sum_{t=2}^{T-1} \tilde{y}'_{t-1}Z_t(Z_t'Z_t)^{-1}Z_t'\tilde{\varepsilon}_t)$ , the second term converges in distribution to a random variable with mean zero. Notice that the second term does not involve higher powers of  $\varepsilon_{i,t}$  (higher than one) and is therefore symmetrically distributed around zero and also uncorrelated with the denominator of  $\hat{r} - r$ ,  $\sum_{t=2}^{T-1} \tilde{y}'_{t-1}Z_t(Z_t'Z_t)^{-1}Z_t'\tilde{y}_{t-1}$ . It follows that

the expectation of second term divided by the denominator is zero. The ratio of the first term and the denominator converges to a negative constant divided by a positive random variable and gives rise to a negative bias.

Finally we remark that we can show in the same way that in fact any Arellano-Bond GMM estimator, i.e. with arbitrary weighting matrix, is biased downwards as for any Arellano-Bond GMM estimator there exists an asymptotically equivalent estimator based on moment conditions involving forward orthogonal deviations.

The distributional results for the Arellano-Bover estimator can be obtained along similar lines after making the following observations. The covariances between the products of the instruments and the regressors, and the products of the instruments and the error terms,  $cov(y_{i,k}\Delta y_{i,l}, (\varepsilon_{i,s} + (1-r)\mu_i)\Delta y_{i,t})$ , are all  $O(1)$ , because  $E(\Delta y_{i,l}\Delta y_{i,t})$ ,  $E(y_{i,k}\Delta y_{i,t})$ ,  $E(y_{i,k}\varepsilon_{i,s})$ ,  $E(\Delta y_{i,l}\varepsilon_{i,s})$ , and  $E(y_{i,k}\mu_i)$  are all  $O(1)$  (the first two are proportional to  $\frac{1}{1+r}$ ). It follows that  $\frac{1}{N}y'_{s-1}Z_s$  and  $N^{-0.5}Z'_t\varepsilon_t$  are asymptotically uncorrelated, where  $Z_t$  contains the instruments for the  $t$ -th equation:  $Z_t = [z_{1,t} \dots z_{N,t}]'$ , with  $z_{i,t} = [\Delta y_{i,1} \dots \Delta y_{i,t-1}]'$ . Finally, since  $\sqrt{N}(\hat{r}_{ARBOV} - r)$  converges in distribution,  $\text{plim}_{N \rightarrow \infty} \hat{r}_{ARBOV} = r$ .

#### B.1.4 Proof of theorem 8 (local-to-unity asymptotics for the system estimator):

When  $T = 3$  the System estimator equals

$$\hat{r}_{SYS} = r + \frac{N^{-0.5}x^1w^{11}e^1 + x^1w^{12}e^2 + N^{-0.5}x^2w^{21}e^1 + x^2w^{22}e^2}{x^1w^{11}x^1 + x^1w^{12}x^2 + x^2w^{21}x^1 + x^2w^{22}x^2} \text{ where}$$

$$x^1 = \frac{1}{N} \sum_{i=1}^N y_{i,2} \Delta y_{i,2},$$

$$x^2 = \frac{1}{N} \sum_{i=1}^N y_{i,1} \Delta y_{i,2},$$

$$e^1 = N^{-0.5} \sum_{i=1}^N (\varepsilon_{i,3} + \mu_i) \Delta y_{i,2},$$

$$e^2 = \frac{1}{N} \sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3}, \text{ and}$$

the  $w^{ij}$  are the elements of the weighting matrix.

When  $T = 3$  the optimal weighting matrix for the System estimator is:  $\begin{bmatrix} \frac{2}{1+r} & -\frac{1}{1+r} \\ -\frac{1}{1+r} & 2b + \frac{2}{1-r^2} \end{bmatrix}^{-1}$

$= \begin{bmatrix} 2(-b + br^2 - 1) \frac{1+r}{4br^2 - r - 4b - 3} & \frac{1}{4br^2 - r - 4b - 3} (-1 + r^2) \\ \frac{1}{4br^2 - r - 4b - 3} (-1 + r^2) & \frac{2}{4br^2 - r - 4b - 3} (-1 + r^2) \end{bmatrix}$ , where  $b = \frac{1}{N} \sum_{i=1}^N \frac{\mu_i^2}{\sigma^2} \geq 0$ , so that

$4br^2 - r - 4b - 3 < -2$ . Therefore,  $\lim_{r \uparrow 1} \begin{bmatrix} \frac{2}{1+r} & -\frac{1}{1+r} \\ -\frac{1}{1+r} & 2b + \frac{2}{1-r^2} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ . So when  $r \uparrow 1$ ,

the Arellano-Bond moment condition receives less and less weight relative to the Arellano-Bover moment condition, and when  $r = 1$ , the former is totally ignored. However, this

analysis is potentially misleading as the elements of the weighting matrix multiply sums that require scaling by different powers of  $N$  in order to converge.

$$\text{When using the sequence } r = 1 - \frac{\lambda}{N}, \begin{bmatrix} w^{11} & w^{12} \\ w^{21} & w^{22} \end{bmatrix} = \begin{bmatrix} O(1) & O(N^{-1}) \\ O(N^{-1}) & O(N^{-1}) \end{bmatrix}.$$

It follows that the denominator converges to the limit of  $x^1 w^{11} x^1$  when  $N \rightarrow \infty$ . After multiplying by  $N^{0.5}$  the numerator converges to the limit of  $x^1 w^{11} e^1$ . So we (can still) conclude that the local-to-unity asymptotic properties of the system estimator are identical to those of the Arellano-Bover estimator.

When  $T > 3$ , the three blocks in the optimal weighting matrix that correspond to the Arellano-Bond moment conditions (one pure block and two cross blocks) still vanish when  $r \uparrow 1$ . This is easily seen as follows. The Arellano-Bond elements  $E(y_{i,k} \Delta \varepsilon_{i,l} y_{i,s} \Delta \varepsilon_{i,t})$  in the covariance matrix of the moment conditions involve terms like  $\frac{1}{1-r^2}$ , which explode when  $r \uparrow 1$ . Moreover, the elements of the inverse of a matrix  $M$  are given by  $\frac{\det M^{ij}}{\det M}$ , where  $M^{ij}$  is the submatrix of  $M$  obtained after deleting its  $i$ -th row and  $j$ -th column. Therefore, only the elements in the Arellano-Bover block of the weighting matrix do not vanish, as only the determinants of the corresponding submatrices of the covariance matrix of the moment conditions (the ‘ $\det M^{ij}$ ’) depend on all the exploding elements, just like the determinant of the covariance matrix of the moment conditions itself.

### B.1.5 Proof of theorem 9:

We will prove that the GMM estimator for  $r$  that is based on the Arellano-Bover moment condition  $E((y_{i,3} - r y_{i,2})(\Delta y_{i,2})) = 0$  is consistent when  $t_0 > -\infty$  is fixed and  $|y_{i,t_0}| < \infty$ .

$$\text{The estimator is given by } \hat{r}_{ABOV} = \frac{\frac{1}{N} \sum_{i=1}^N y_{i,3} \Delta y_{i,2}}{\frac{1}{N} \sum_{i=1}^N y_{i,2} \Delta y_{i,2}}.$$

$$\text{Then } \hat{r}_{ABOV} - 1 = \frac{\frac{1}{N} \sum_{i=1}^N y_{i,3} \Delta y_{i,2}}{\frac{1}{N} \sum_{i=1}^N y_{i,2} \Delta y_{i,2}} - 1 = \frac{\frac{1}{N} \sum_{i=1}^N \varepsilon_{i,3} \varepsilon_{i,2}}{\frac{1}{N} \sum_{i=1}^N y_{i,2} \varepsilon_{i,2}}.$$

We condition on  $y_{i,t_0}$ ,  $i = 1, \dots, N$ . Then we can decompose the denominator  $\frac{1}{N} \sum_{i=1}^N y_{i,2} \varepsilon_{i,2}$  as  $\frac{1}{N} \sum_{i=1}^N \varepsilon_{i,2}^2 + \frac{1}{N} \sum_{i=1}^N \sum_{s=t_0+1}^1 \varepsilon_{i,s} \varepsilon_{i,2} + \frac{1}{N} \sum_{i=1}^N y_{i,t_0} \varepsilon_{i,2}$ .

We have  $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \varepsilon_{i,3} \varepsilon_{i,2} = 0$ ,  $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \varepsilon_{i,2}^2 = \sigma^2$ ,  $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N y_{i,t_0} \varepsilon_{i,2} = 0$ , and  $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{s=t_0+1}^1 \varepsilon_{i,s} \varepsilon_{i,2} = 0$ .

It follows straightforwardly that  $\text{plim}_{N \rightarrow \infty} \hat{r}_{ABOV} = 1$ .  $\square$

## B.2 Proofs of the results in section 4.2

### B.2.1 Proof of theorem 11 (weak instruments):

Let  $m_1(r) = y_{i,1}(\Delta y_{i,t} - r\Delta y_{i,t-1})$ ;  $m_2(r) = (y_{i,t} - ry_{i,t-1})^2 - (y_{i,2} - ry_{i,1})^2$ ; and  $m_3(r) = (y_{i,t} - ry_{i,t-1})(y_{i,s} - ry_{i,s-1}) - (y_{i,3} - ry_{i,2})(y_{i,2} - ry_{i,1})$ ; and  $m_4(r) = \Delta y_{i,t} - r\Delta y_{i,t-1}$ .

Noticing that  $E(y_{i,t}) = rE(y_{i,t-1}) + (1-r)\mu_i$ ,  $t = 2, \dots, T$ , and  $E(y_{i,2}) = ry_{i,1} + (1-r)\mu_i$ , we obtain the following results:

$$\frac{dm_1}{dr} = -y_{i,1}\Delta y_{i,t-1}, \text{ and } E\left(\frac{dm_1}{dr}\right) = (1-r)E(y_{i,1}(y_{i,t-2} - \mu_i)).$$

Remember that the  $\mu_i$  are fixed. Then, conditional on  $y_{i,1}$ ,  $E(y_{i,1}(y_{i,t-2} - \mu_i)) = r^{t-3}y_{i,1} \times (y_{i,1} - \mu_i) \neq 0$ , in general, when  $|r| < 1$ , and  $\lim_{r \uparrow 1} E\left(\frac{dm_1}{dr}\right) = 0$ . [Under stationarity  $E\left(\frac{dm_1}{dr}\right) = E(-y_{i,1}\Delta y_{i,t-1}) = \frac{r^{t-3}}{1+r}\sigma^2$ .]

$$\begin{aligned} \frac{dm_2}{dr} &= -2y_{i,t-1}(y_{i,t} - ry_{i,t-1}) + 2y_{i,1}(y_{i,2} - ry_{i,1}) = \\ &-2y_{i,t-1}(1-r)\mu_i + 2y_{i,1}(1-r)\mu_i - 2y_{i,t-1}\varepsilon_t + 2y_{i,1}\varepsilon_2, \end{aligned}$$

$$E\left(\frac{dm_2}{dr}\right) = E[-2y_{i,t-1}(1-r)\mu_i + 2y_{i,1}(1-r)\mu_i], \text{ and } \lim_{r \uparrow 1} E\left(\frac{dm_2}{dr}\right) = 0.$$

$$\begin{aligned} \frac{dm_3}{dr} &= -y_{i,t-1}(y_{i,s} - ry_{i,s-1}) - y_{i,s-1}(y_{i,t} - ry_{i,t-1}) + y_{i,2}(y_{i,2} - ry_{i,1}) + y_{i,1}(y_{i,3} - ry_{i,2}) = \\ &-y_{i,t-1}(1-r)\mu_i - y_{i,s-1}(1-r)\mu_i + y_{i,2}(1-r)\mu_i + y_{i,1}(1-r)\mu_i - y_{i,t-1}\varepsilon_s - y_{i,s-1}\varepsilon_t + y_{i,2}\varepsilon_2 + y_{i,1}\varepsilon_3, \\ E\left(\frac{dm_3}{dr}\right) &= E[-y_{i,t-1}(1-r)\mu_i - y_{i,s-1}(1-r)\mu_i + y_{i,2}(1-r)\mu_i + y_{i,1}(1-r)\mu_i] - r^{t-s-1}\sigma^2 + \sigma^2, \end{aligned}$$

and

$$\lim_{r \uparrow 1} E\left(\frac{dm_3}{dr}\right) = 0.$$

$$\frac{dm_4}{dr} = -\Delta y_{i,t-1}, \text{ and } \lim_{r \uparrow 1} E\left(\frac{dm_4}{dr}\right) = 0, \text{ because } \lim_{r \uparrow 1} E(\Delta y_{i,1}) = \lim_{r \uparrow 1} (r-1)(y_{i,1} - \mu_i) = 0$$

We conclude that all the moment conditions corresponding to the conditional ARFE(1) model are weak when  $r$  is close to one. Notice that  $E[m_2(r)] = 0$  and  $E[m_3(r)] = 0$  are also weak under stationarity.

### B.2.2 Proof of theorem 12 (local-to-zero asymptotics):

The conditional model reads:

$$y_i = ry_{i-1} + \check{\mu}_i \iota + \varepsilon_i, \text{ where } \check{\mu}_i = (1-r)\mu_i, \text{ and } \varepsilon_i | \check{\mu}_i, y_{i,1} \sim (0, \sigma^2 I_{T-1}) \text{ (ind. dist.)}.$$

When  $T = 3$  we have one Arellano-Bond moment condition:

$$E[y_{i,1}(\Delta y_{i,3} - r\Delta y_{i,2})] = E(y_{i,1}\Delta \varepsilon_{i,3}) = 0.$$

We have the following results:

$$E(y_{i,1}\Delta y_{i,2}) = y_{i,1}(r-1)(y_{i,1} - \mu_i),$$

$$\text{Var}(y_{i,1}\Delta y_{i,2}) = \text{Var}(y_{i,1}\varepsilon_{i,2}) = \sigma^2 y_{i,1}^2,$$

$$E(y_{i,1}\Delta\varepsilon_{i,3}) = 0,$$

$$\text{Var}(y_{i,1}\Delta\varepsilon_{i,3}) = 2\sigma^2 y_{i,1}^2, \text{ and}$$

$$\text{Cov}(y_{i,1}\Delta\varepsilon_{i,3}, y_{i,1}\Delta y_{i,2}) = -\sigma^2 y_{i,1}^2.$$

Let  $r = 1 - \lambda N^{-0.5}$  (cf Staiger and Stock, 1997) and let us define

$$Z_1 = N^{-0.5} \sum_{i=1}^N y_{i,1} \Delta\varepsilon_{i,3}, \text{ and}$$

$$Z_2 = N^{-0.5} \sum_{i=1}^N y_{i,1} \Delta y_{i,2}.$$

$$\text{Then with } r = 1 - \lambda N^{-0.5} \text{ } Z_1 \xrightarrow{a} N(0, 2\sigma^2 \frac{1}{N} \sum_{i=1}^N y_{i,1}^2),$$

$$Z_2 \xrightarrow{a} N(-\lambda \frac{1}{N} \sum_{i=1}^N y_{i,1}(y_{i,1} - \mu_i), \sigma^2 \frac{1}{N} \sum_{i=1}^N y_{i,1}^2),$$

where  $\xrightarrow{a}$  stands for ‘‘has large sample distribution’’.

$$\text{Cov}(Z_1, Z_2) = -\sigma^2 \frac{1}{N} \sum_{i=1}^N y_{i,1}^2.$$

It follows that  $\hat{r}_{ARBOND} \xrightarrow{a} r + \frac{\check{Z}_1}{\check{Z}_2}$ , where

$$\begin{bmatrix} \check{Z}_1 \\ \check{Z}_2 \end{bmatrix} \sim N \left[ \begin{pmatrix} 0 \\ -\lambda \left( \frac{1}{N} \sum_{i=1}^N y_{i,1}(y_{i,1} - \mu_i) \right) \end{pmatrix}, \sigma^2 \left( \frac{1}{N} \sum_{i=1}^N y_{i,1}^2 \right) \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right].$$

When  $E(y_{i,1}) = \mu_i$ ,  $E|y_{i,1}|^{2+\delta} < \infty$ , and assuming that the  $y'_{i,1}$ s are independent,  $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N y_{i,1}(y_{i,1} - \mu_i) = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,1} - \mu_i)^2 < \infty$ .

### B.2.3 Proof of theorem 13 (unit root):

When  $r = 1$ , we have  $y_{i,t} = y_{i,t-1} + \varepsilon_{i,t}$ .

Therefore we obtain

$$\frac{dm_1}{dr} = -y_{i,1}\Delta y_{i,t-1} = -y_{i,1}\varepsilon_{i,t-1}, \text{ and } E\left(\frac{dm_1}{dr}\right) = 0,$$

$$\frac{dm_2}{dr} = -2y_{i,t-1}(y_{i,t} - y_{i,t-1}) + 2y_{i,1}(y_{i,2} - y_{i,1}) = -2y_{i,t-1}\varepsilon_t + 2y_{i,1}\varepsilon_2, \text{ and } E\left(\frac{dm_2}{dr}\right) = 0,$$

$$\begin{aligned} \frac{dm_3}{dr} &= -y_{i,t-1}(y_{i,s} - y_{i,s-1}) - y_{i,s-1}(y_{i,t} - y_{i,t-1}) + y_{i,2}(y_{i,2} - y_{i,1}) + y_{i,1}(y_{i,3} - y_{i,2}) = \\ &= -y_{i,t-1}\varepsilon_s - y_{i,s-1}\varepsilon_t + y_{i,2}\varepsilon_2 + y_{i,1}\varepsilon_3, \text{ and } E\left(\frac{dm_3}{dr}\right) = -\sigma^2 + \sigma^2 = 0, \end{aligned}$$

$$\frac{dm_4}{dr} = -\Delta y_{i,t-1} = -\varepsilon_{i,t-1}, \text{ and } E\left(\frac{dm_4}{dr}\right) = 0.$$

We conclude that  $r = 1$  is not identified by the moment conditions corresponding to the conditional ARFE(1) model.

## References

- [1] Abowd, J., and D. Card, 1989, On the covariance structure of earnings and hours changes, *Econometrica* 57, 411-445.
- [2] Ahn, S.C., and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, *Journal of econometrics* 68, 5-28.
- [3] Ahn, S.C., and P. Schmidt, 1997, Efficient estimation of dynamic panel data models: alternative assumptions and simplified estimation, *Journal of econometrics* 76, 309-321.
- [4] Alonso-Borrego, C., and M. Arellano, 1996, Symmetrically normalised instrumental variable estimation using panel data, *Journal of Business & Economic Statistics* 17, 36-49.
- [5] Arellano, M., and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of economic studies* 58, 277-297.
- [6] Arellano, M., and O. Bover, 1995, Another look at the instrumental variable estimation of error-components models, *Journal of econometrics* 68, 29-51.
- [7] Blundell, R.W., and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, *Journal of econometrics* 87, 115-143.
- [8] Blundell, R.W., and S. Bond, 1999, GMM estimation with persistent panel data: an application to production function estimation, Working paper series No. W99/4, The Institute of Fiscal Studies, London.
- [9] Bound, J., D.A. Jaeger, and R.M. Baker, 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, *Journal of the American statistical association* 90, 443-450.
- [10] Breitung, J., and W. Meyer, 1994, Testing for unit roots using panel data: are wages on different bargaining levels cointegrated?, *Applied economics* 26, 353-361.



- [11] Breusch, T., H. Qian, P. Schmidt, and D. Wyhowski, 1999, Redundancy of moment conditions, *Journal of econometrics* 91, 89-111.
- [12] Chamberlain, G., 1982, Multivariate regression models for panel data, *Journal of econometrics* 18, 5-46.
- [13] Chamberlain, G., 1984, Panel data, in *Handbook of econometrics*, edited by Z. Griliches and M. Intrilligator, North Holland, Amsterdam.
- [14] Crépon, B., F. Kramarz, and A. Trognon, 1997, Parameters of interest, nuisance parameters, and orthogonality conditions, *Journal of econometrics* 82, 135-156.
- [15] Fieller, E.C., 1932, The distribution of an index in a normal bivariate population, *Biometrika* 24, 428-440.
- [16] Griliches, Z., and J. Mairesse, 1998, Production functions: the search for identification, in S. Strom (ed.), *Essays in honour of Ragnar Frisch*, *Econometric Society Monograph Series*, Cambridge University Press, Cambridge, U.K. .
- [17] Hahn, J., 1999, How informative is the initial condition in the dynamic panel model with fixed effects?, *Journal of econometrics* 93, 309-326.
- [18] Harris, R.D.F., and E. Tzavalis, 1999, Inference for unit roots in dynamic panels where the time dimension is fixed, *Journal of econometrics* 91, 201-226.
- [19] Kiviet, J.F., 1995, On the bias, inconsistency, and efficiency of various estimators in dynamic panel data models, *Journal of econometrics* 68, 53-78.
- [20] Kruiniger, H., 1996, Time-to-build and time-to-innovate in a model of interrelated factor demand with adjustment costs, mimeo, INSEE-CREST, Malakoff, France.
- [21] Kruiniger, H., 1998a, Conditional maximum likelihood estimation of dynamic panel data models, *University College London Economics Discussion Paper* 98-04.
- [22] Kruiniger, H., 1998b, CML and GMM based unit root tests for dynamic panel data models with fixed effects, *Agricultural Economic Research Institute LEI-DLO*, The Hague, mimeographed.

- [23] Kruiniger, H., 2000a, ML and GMM estimation of dynamic panel data models with fixed effects, Queen Mary and Westfield College, University of London, revised and mimeographed.
- [24] Kruiniger, H., 2000b, A bias-corrected Minimum Distance estimator for conditional dynamic panel data models with fixed effects, Queen Mary and Westfield College, University of London, mimeographed.
- [25] MaCurdy, T., 1981, Multiple time series models applied to panel data, NBER working paper no. W0646, Cambridge, Massachusetts.
- [26] MaCurdy, T., 1982, The use of time series processes to model the time structure of earnings in a longitudinal data analysis, *Journal of econometrics* 18, 83-114.
- [27] Maddala, G.S., and J. Jeong, 1992, On the exact small sample distribution of the instrumental variable estimator, *Econometrica* 60, 181-183.
- [28] Nelson, C.R., and R. Startz, 1990a, The distribution of the instrumental variable estimator and its t-ratio when the instrument is a poor one, *Journal of business* 63, 5125-5140.
- [29] Nelson, C.R., and R. Startz, 1990b, Some further results on the exact small sample properties of the instrumental variable estimator, *Econometrica* 49, 1417-1426.
- [30] Neyman, J., and E.L. Scott, 1948, Consistent estimates based on partially consistent observations, *Econometrica* 16, 1-32.
- [31] Nickell, S., 1981, Biases in dynamic models with fixed effects, *Econometrica* 49, 1417-1426.
- [32] Staiger, D., and J.H. Stock, 1997, Instrumental variables regression with weak instruments, *Econometrica* 65, 557-586.

**This working paper has been produced by  
the Department of Economics at  
Queen Mary, University of London**

**Copyright © 2000 Hugo Kruiniger  
All rights reserved.**

**Department of Economics  
Queen Mary, University of London  
Mile End Road  
London E1 4NS  
Tel: +44 (0)20 7882 5096 or Fax: +44 (0)20 8983 3580  
Email: [j.conner@qmw.ac.uk](mailto:j.conner@qmw.ac.uk)  
Website: [www.econ.qmw.ac.uk/papers/wp.htm](http://www.econ.qmw.ac.uk/papers/wp.htm)**