Moment restrictions and identification in linear dynamic panel data models

By

Tue Gørgens
Chirok Han
Sen Xue

ANU Working Papers in Economics and Econometrics
# 633

March 2016  JEL: C230

ISBN: 0 86831 633 4
Moment restrictions and identification in linear dynamic panel data models

Tue Gørgens∗ Chirok Han† Sen Xue‡

2016-03-03

Abstract: This paper investigates the relationship between moment restrictions and identification in simple linear AR(1) dynamic panel data models with fixed effects under standard minimal assumptions. The number of time periods is assumed to be small. The assumptions imply linear and quadratic moment restrictions which can be used for GMM estimation. The paper makes three points. First, contrary to common belief, the linear moment restrictions may fail to identify the autoregressive parameter even when it is known to be less than 1. Second, the quadratic moment restrictions provide full or partial identification in many of the cases where the linear moment restrictions do not. Third, the first moment restrictions can also be important for identification. Practical implications of the findings are illustrated using Monte Carlo simulations.

Keywords: Dynamic panel data models, fixed effects, identification, generalized method of moments, Arellano-Bond estimator.

JEL classification codes: C230.

∗The Australian National University. E-mail: tue.gorgens@anu.edu.au.
†Korea University. E-mail: chirokhan@korea.ac.kr.
‡The Australian National University. E-mail: sen.xue@anu.edu.au.
1 Introduction

In this paper we consider moment restrictions and identification of the autoregressive parameter in simple linear AR(1) dynamic panel data models with fixed effects. We consider the case where the number of individuals is large and the number of time periods is small. We impose only the standard minimal assumptions that the idiosyncratic errors are serially uncorrelated and uncorrelated with the initial data values and the fixed effects. The assumptions imply a number of linear and quadratic moment restrictions on the covariance matrix of the data, which can be used to identify and estimate the autoregressive parameter (see e.g. Arellano and Bond, 1991; Ahn and Schmidt, 1995).

There are three main points in this paper. First, we demonstrate that there are cases not previously recognized in the literature where the linear Arellano-Bond moment restrictions do not identify the autoregressive parameter. It is common to assume that the autoregressive parameter is different from 1 in order to rule out an obvious “random walk” case of unidentification. Our results show that this assumption is insufficient. It does not ensure identification. In general identification failure can arise for any value of the autoregressive parameter. Even if trivial cases with no time variation in the data are ruled out, identification of the autoregressive parameter can fail for any value in the interval $[0, 1]$. The linear moment restrictions underpin the well-known “difference GMM” estimator. The difference GMM estimator is inconsistent when the autoregressive parameter is not identified.

Second, we show that the quadratic moment restrictions provide identification in many of the cases where the linear moment restrictions do not. The previous literature has focused on the potential efficiency gains in utilizing the quadratic moment restrictions. Our results show that the quadratic moment restrictions can also play a more fundamental role by providing the only source of identification. Moreover, even when both the linear and the quadratic moment restrictions fail, the latter provide partial identification as long as there is some time variation in the data. Partial identification comes in the form of two distinct candidates, one of which equals the true parameter. We characterize identification both in terms of the moments of the observed data and in terms of the underlying parameters.

Third, we show that restrictions on the means (as opposed to the second moments) of
the data can be important for identification as well as for efficiency.

The simple linear AR(1) model has a long history in econometrics, with important early work by Balestra and Nerlove (1966), Anderson and Hsiao (1981, 1982), Holtz-Eakin, Newey, and Rosen (1988), and Arellano and Bond (1991). Ahn and Schmidt (1995) presented the complete set of second moment restrictions available under a range of assumptions. Han and Kim (2014) argued that a constant term should be included among the instruments; this is effectively the same as utilizing first moment restrictions. Both Ahn and Schmidt (1995) and Han and Kim (2014) were concerned mainly with efficient estimation and did not discuss the details of identification.

It appears that the issue of identification has not previously been studied systematically. Blundell and Bond (1998) showed that identification problems arise for the difference GMM estimator in the limit as the autoregressive parameter approaches 1 and the variance of the fixed effects approaches 0. Recently, Bun and Kleibergen (2013) also studied identification (and distributions of estimators and test statistics) in the limit under similar conditions. They considered several GMM estimators and found that identification may fail in the limit depending on the rate at which the autoregressive parameter approaches 1. The present paper investigates identification for all logically possible values of the underlying parameters.

The paper is organized as follows. Section 2 presents the model and the standard minimal assumptions. Section 3 considers the identification provided by the familiar linear (Arellano-Bond) moment restrictions. Section 4 considers the identification provided by the quadratic (Ahn-Schmidt) moment restrictions when the linear moment restrictions do not identify the parameter of interest. Section 5 considers first moment restrictions. Section 6 discusses supplementary assumptions and moment restrictions sometimes considered in the literature. Section 7 presents simulation results to illustrate practical implications of identification failure. Section 8 concludes. Appendix A contains proofs of theorems, and Appendix B provides general results for the so-called “stationarity” assumption and the “level” moment restrictions.
2 The simple AR(1) dynamic panel data model

Suppose we have two-dimensional panel data. Following common practice, we shall refer to the first dimension as individuals and the second as time periods. For \( i = 1, \ldots, n \) individuals and \( t = 0, \ldots, T \) times, let \( y_{it} \) be scalar random variables which are observed and available for analysis.\(^1\) Note that \( y_{i0} \) is observed.

**Assumption 1** The random variables \( \{(y_{i0}, \ldots, y_{iT}) : i = 1, \ldots, n\} \) are independent across individuals, have finite means, and satisfy

\[
y_{it} = y_{i(t-1)} + c_i + v_{it}, \quad t = 1, \ldots, T,
\]

with \( E(v_{it}) = 0 \) for \( t = 1, \ldots, T \).

In Assumption 1, \( \alpha_0 \) is an unknown parameter to be estimated, and the sum \( c_i + v_{it} \) is an unobserved term decomposed into an individual-specific random variable, \( c_i \), which is constant over time, and an individual- and time-specific term, \( v_{it} \). The model does not have an explicit constant term; this is subsumed into \( c_i \).

It is conventional in the theoretical literature to assume \( E(c_i) = 0 \), sometimes accompanied by a remark that a nonzero mean can be handled by including a constant term in the model (see e.g. Ahn and Schmidt, 1995; Blundell and Bond, 1998; Alvarez and Arellano, 2003). We eschew that assumption in this paper in order to discuss certain implications. We also consider models with time effects in Section 5.

Additional assumptions are usually imposed on this distribution in order to identify and estimate the main parameter of interest, \( \alpha_0 \). The standard minimal set is stated as Assumption 2 (see also Ahn and Schmidt, 1995).

---

\(^1\)For simplicity, the expression “\( i = 1, \ldots, n \)” is suppressed in the remainder of this paper.
Assumption 2 The random variables in Assumption 1 have finite variances and satisfy

\[ E(v_{it}y_{i0}) = 0, \quad t = 1, \ldots, T; \]  
\[ E(v_{it}c_i) = 0, \quad t = 1, \ldots, T; \]  
\[ E(v_{is}v_{it}) = 0, \quad s = 1, \ldots, t - 1, \quad t = 2, \ldots, T. \]

In general, there are potentially \( T+1 \) distinct first moments and \((T+1)(T+2)/2\) distinct second moments of the observed variables. However, under Assumptions 1 and 2, there are 3 parameters which determine the first moments of the data, namely \( \alpha_0, E(c_i), \) and \( E(y_{i0}) \), and there are \( 4 + T \) parameters which determine the second moments of the joint distribution of the data, namely \( \alpha_0, E(y_{i0}^2), E(c_i^2), E(y_{i0}c_i), \) and \( E(v_{it}^2) \) for \( t = 1, \ldots, T. \)

It is customary in the literature to assume that \( |\alpha_0| < 1 \) (see e.g. Arellano and Bond, 1991; Ahn and Schmidt, 1995; Blundell and Bond, 1998; Alvarez and Arellano, 2003). In this paper we avoid such assumptions and consider identification of \( \alpha_0 \) for all possible true parameter values for a complete characterization.

Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bond (1991) considered estimating \( \alpha_0 \) using the following \( T(T-1)/2 \) linear moment restrictions:

\[ E(y_{is}(\Delta y_{it} - \Delta y_{it-1} \alpha)) = 0, \quad s = 0, \ldots, t - 2, \quad t = 2, \ldots, T. \]

For convenience, we refer to (5) as the linear AB moment restrictions. Ahn and Schmidt (1995) derived the following \( T - 2 \) quadratic moment restrictions:

\[ E((y_{iT} - y_{iT-1} \alpha)(\Delta y_{it} - \Delta y_{it-1} \alpha)) = 0, \quad t = 2, \ldots, T - 1. \]

We refer to (6) as the quadratic AS moment restrictions. Ahn and Schmidt (1995) showed that under Assumption 2 there are exactly \( T(T-1)/2 + (T-2) \) restrictions on the second moments of the data and that these can be represented in several ways. They also showed

\(^2\text{For identification analysis, it does not matter whether we work with second moments about the origin or about the mean. In practice, due to the nonlinear relationship between them, GMM estimation results will not be numerically identical.}\)
that the set given in (5) and (6) is the representation of these restrictions which maximizes
the number of linear restrictions, and they investigated the efficiency gains arising from (6),
among other matters.

As mentioned, there is an additional set of restrictions, based on the first moments, which
can also play an important role in identifying $\alpha_0$. We discuss these restrictions in Section 5.

Assumption 2 implies that (5) and (6) are satisfied for $\alpha = \alpha_0$. The question of identifica-
tion is whether or not $\alpha = \alpha_0$ is a unique solution. In the remainder of the paper, we show
that identification cannot be taken for granted even under the restriction that $|\alpha_0| < 1$.
Authors who assume $|\alpha_0| < 1$ in small-$T$ settings presumably hope that this will ensure
identification of $\alpha_0$. We show that assuming $|\alpha_0| < 1$ is ineffective when only the linear AB
moment restrictions are used. Assuming $|\alpha_0| < 1$ can be effective if both the linear and the
quadratic moment restrictions are used, but as yet very few authors have considered the
quadratic moment restrictions.

3 The linear AB moment restrictions

Following the seminal papers by Holtz-Eakin, Newey, and Rosen (1988) and Arellano and
Bond (1991), many studies have estimated $\alpha_0$ using the so-called difference GMM estimator,
which is based on the linear AB moment restrictions. Obviously, this estimator is convenient
since it can be computed by matrix inversion and iteration is not required. In this section,
we show that identification failure is possible for any value of $\alpha_0$.

We begin by characterizing identification failure in terms of the second moments of the
data. From (5), it is immediately clear that the linear AB moment restrictions fail to identify
$\alpha_0$ if and only if

$$
E(y_{is}\Delta y_{it-1}) = 0, \quad s = 0, \ldots, t - 2, \quad t = 2, \ldots, T. \quad (7)
$$

If (7) holds, then any value of $\alpha$ will satisfy (5).

For a visual illustration of when identification of $\alpha_0$ fails, note that Assumptions 1 and 2
imply $E(y_{is}\Delta y_{it-1}) = \alpha_0 E(y_{is}\Delta y_{it-2}) = \cdots = \alpha_0^{t-2-s} E(y_{is}\Delta y_{it+1})$ for $s = 0, \ldots, t - 2$ and $t =$
Thus, necessary and sufficient conditions for identification failure by the linear AB moment restrictions are that
\[ E(y_{it} \Delta y_{it+1}) = 0 \] for \( t = 0, \ldots, T - 2 \). These conditions can be written equivalently
\[ E(y_{it}^2) = E(y_{it} y_{is}), \quad s = t, \ldots, T, \quad t = 0, \ldots, T - 2. \] (8)

This shows that identification failure is associated with a distinct pattern in the matrix of second moments of the data. If the rows and columns in this matrix is numbered from 0 to \( T \) and the typical element is \( E(y_{it} y_{is}) \), where \( t \) represents the row and \( s \) the column, then (8) implies that the entries are identical across columns \( t \) to \( T - 1 \) for each row \( t = 0, \ldots, T - 2 \), while the lower right \( 2 \times 2 \) submatrix is unrestricted. Figure 1 illustrates this pattern.

The equations in (7) or (8) provide a characterization of identification failure in terms of properties of the observed data. Identification failure can also be described in terms of the underlying data generating processes. It is well known that identification by linear AB moment restrictions fails in the “random walk” case, where \( \alpha_0 = 1 \) and \( E(c_i^2) = 0 \); that is, when \( y_{it} = y_{it-1} + v_{it} \) for \( t = 1, \ldots, T \). This particular case is excluded by assuming that \( |\alpha_0| < 1 \). However, it turns out that there are many other cases, including cases with \( |\alpha_0| < 1 \), where identification fails. The following example is illustrative.

**Example 1** Let \( \alpha_0 \) be an arbitrary value between 0 and 1. Let \( y_{i0}, w_i, v_{i1}, v_{i2}, \) and \( v_{i3} \) be mutually independent random variables with zero mean. Let \( y_{it} \) be generated by \( y_{it} = \alpha_0 y_{it-1} + c_i + v_{it} \) for \( t = 1, 2, 3 \), where \( c_i = (1 - \alpha_0)y_{i0} + w_i \). Then

\[ y_{i1} = \alpha_0 y_{i0} + c_i + v_{i1} = y_{i0} + w_i + v_{i1}, \] (9)

\[ y_{i2} = \alpha_0 y_{i1} + c_i + v_{i2} = y_{i0} + (1 + \alpha_0)w_i + v_{i2} + \alpha_0 v_{i3}, \] (10)

and thus \( \Delta y_{i1} = w_i + v_{i1} \) and \( \Delta y_{i2} = \alpha_0 w_i + v_{i2} - (1 - \alpha_0)v_{i3} \). From these, we have

---

3 The adopted convention is \( \alpha^0 = 1 \) if \( \alpha = 0 \).

4 The equation for \( s = T \) follows from \( E(y_{it} \Delta y_{iT}) = \alpha_0 E(y_{it} \Delta y_{iT-1}) = 0 \) for \( t = 0, \ldots, T - 2 \), where the last equality follows directly from the failure of the linear AB moment restrictions to identify \( \alpha_0 \). In Figure 1, row \( T \) and column \( T \) play no direct role in the identification of \( \alpha_0 \) by the linear AB moment restrictions.
\[ E(y_{i0} \Delta y_{i1}) = 0 \text{ and } E(y_{i0} \Delta y_{i2}) = 0. \] If \( \alpha_0 E(w_{i1}^2) = (1 - \alpha_0) E(v_{i1}^2) \), then we also have
\[ E(y_{i1} \Delta y_{i2}) = \alpha_0 E(w_{i1}^2) - (1 - \alpha_0) E(v_{i1}^2) = 0. \quad (11) \]

That is, (5) is satisfied by any \( \alpha \in \mathbb{R} \). Note that any value of \( \alpha_0 \) between 0 and 1 is compatible with the condition \( \alpha_0 E(w_{i1}^2) = (1 - \alpha_0) E(v_{i1}^2) \).

Example 1 is interesting because the linear AB moment restrictions fail to identify the true parameter even though the data generating process satisfies all the standard assumptions and also the true parameter is less than unity. Remarkably, combining the standard assumptions and \( |\alpha_0| < 1 \) is not sufficient for identification.

To understand precisely when the linear AB moment restrictions do not identify \( \alpha_0 \), Theorem 1 fully characterizes identification for all possible data generating processes (DGPs).

**Theorem 1** Suppose Assumptions 1 and 2 hold. The linear AB moment restrictions fail to identify \( \alpha_0 \) if and only if

\[ (\alpha_0 - 1) E(y_{i0}^2) + E(y_{i0} c_i) = 0 \quad (12) \]

and, for \( T \geq 3 \),

\[ \alpha_0' \Psi + (\alpha_0 - 1) E(v_{i1}^2) = 0, \quad t = 1, \ldots, T - 2, \quad (13) \]

where \( \Psi = (\alpha_0 - 1) E(c_i y_{i0}) + E(c_i^2) \).

If \( T = 2 \), any value of \( \alpha_0 \) is compatible with (12).

If \( T \geq 3 \), \( \Psi = 0 \) and \( E(v_{i1}^2) = 0 \) for all \( t = 1, \ldots, T - 2 \), then any value of \( \alpha_0 \) is compatible with (12) and (13).

If \( T \geq 3 \) and either \( \Psi \neq 0 \) or \( E(v_{i1}^2) \neq 0 \) for some \( t = 1, \ldots, T - 2 \), then only values of \( \alpha_0 \) in \([0, 1]\) are compatible with (12) and (13).

The proof can be found in Appendix A.1. Theorem 1 shows that for any value of \( \alpha_0 \) there are data generating processes such that the linear AB moments do not provide identification. Some of these are not particularly interesting, because the variance of the idiosyncratic error
terms are zero, so \( v_{it} = 0 \) for all individuals and time periods. Nevertheless, it is not difficult to construct nontrivial examples of identification failure where all variances and covariances are nonzero. Example 1 is one. It is easy to extend Example 1 to \( T > 3 \).

One way to understand Theorem 1 is to use the “random walk” case as a reference point. Suppose the true parameters satisfy (12) and (13), so the linear AB moment restrictions do not identify \( \alpha_0 \). One possibility is the case of \( n \) independent random walks, which arises when \( \alpha_0 = 1 \) and \( E(c_i^2) = 0 \). By (5), the linear AB moment restrictions are satisfied for \( \alpha = 1 \) when \( E(y_{is}(\Delta y_{it} - \Delta y_{it-1})) = 0 \) for \( s = 0, \ldots, t - 2 \) and \( t = 2, \ldots, T \). If identification fails and \( \alpha_0 \neq 1 \), then (13) implies \( E(v_{it}^2) = \alpha_0^2(1 - \alpha_0)^{-1}\Psi \) for \( t = 1, \ldots, T - 2 \). That is, the variances of the idiosyncratic error terms constitute a geometric sequence. By Theorem 1, from the perspective of the linear AB moment restrictions, the latter data generating process cannot be distinguished from the random walk case. Define \( \tilde{v}_{it} = \Delta y_{it} \) for \( t = 1, \ldots, T \). Identification failure implies that we must have \( E(y_{is}\Delta \tilde{v}_{it}) = 0 \) for all \( s = 0, \ldots, t - 2 \) and \( t = 2, \ldots, T \). Therefore, the true data generating process looks the same as the random walk case \( y_{it} = y_{it-1} + \tilde{v}_{it} \) for \( t = 1, \ldots, T \).

If the linear AB moment restrictions fail to identify \( \alpha_0 \), then any value of \( \alpha \) is a candidate solution. Theorem 1 shows that in some cases only values of \( \alpha_0 \) between 0 and 1 are compatible with the data. That is, some of the solutions to the linear AB moment restrictions may be incompatible with the data generating process. The difference GMM estimator cannot benefit from this insight, but we show in Section 4 that it is possible to exploit these bounds on \( \alpha_0 \) when identification and estimation is based on both the linear AB and the quadratic AS moment restrictions.

4 The quadratic AS moment restrictions

Are the standard minimal assumptions strong enough to identify \( \alpha_0 \)? Having established the limitations of identification by the linear AB moment restrictions, the natural question is whether the quadratic AS moment restrictions identify \( \alpha_0 \) when the linear do not. To the best of our knowledge, when discussing quadratic moment restrictions the focus has been
on efficiency gains, not identification (see e.g. Ahn and Schmidt, 1995). (Of course, the two aspects are closely related.) In this section, we show that the quadratic AS moment restrictions often provide identification when the linear AB moment restrictions do not identify \( \alpha_0 \).

Written out, the quadratic AS moment restrictions are

\[
E((y_{iT} - y_{iT-1})\alpha)(\Delta y_{it} - \Delta y_{it-1}\alpha)) = E(y_{iT}\Delta y_{it}) - (E(y_{iT-1}\Delta y_{it}) + E(y_{iT}\Delta y_{it-1}))\alpha + E(y_{iT-1}\Delta y_{it-1})\alpha^2 = 0, \\
t = 2, \ldots, T - 1. \quad (14)
\]

Since the equations in (14) are polynomials of second degree or less, the information they can provide about \( \alpha_0 \) falls in one of three categories: either an equation provides unique identification of \( \alpha_0 \), partial identification in the form of two roots one of which must be \( \alpha_0 \), or it provides no information about \( \alpha_0 \) at all (i.e. any value of \( \alpha \) is a solution).

When the linear AB moment restrictions fail to identify \( \alpha_0 \) so (8) holds, the quadratic AS moment restrictions can be simplified to

\[
E(y_{it}^2) - E(y_{it-1}^2) - (E(y_{it}^2) - E(y_{it-2}^2))\alpha + (E(y_{it-1}^2) - E(y_{it-2}^2))\alpha^2 = 0, \\
t = 2, \ldots, T - 2, \quad (15a)
\]

\[
E(y_{iT-1}\Delta y_{iT}) + E(y_{iT-1}^2) - E(y_{iT-2}^2) - (E(y_{iT-1}^2) - E(y_{iT-3}^2))\alpha + (E(y_{iT-2}^2) - E(y_{iT-3}^2))\alpha^2 = 0, \quad (15b)
\]

where (15a) is empty if \( T = 3 \). Note that (15b) deviates from the pattern in (15a) unless \( E(y_{iT-1}\Delta y_{iT}) = 0 \).

The equations in (15) show that the information about \( \alpha_0 \) in the quadratic AS moment restrictions when the linear AB moment restrictions fail to identify \( \alpha_0 \) is related to the time pattern in the second moments \( E(y_{it}^2) \) for \( t = 0, \ldots, T - 1 \). This is in accordance with the illustration in Figure 1, where any information about \( \alpha_0 \) in the second moments of the
data not utilized by the linear AB moment restrictions must come from comparisons across
different-shaded bands, or possibly from the lower right $2 \times 2$ white submatrix.

We provide two theorems about the identification of $\alpha_0$ by the linear AB and quadratic
AS moment restrictions. Theorem 2 describes identification in terms of the second moments
of the data. Theorem 3 characterizes identification issues in terms of the underlying DGPs.
There are no quadratic AS moment restrictions when $T = 2$, and the theorems cover $T \geq 3$.
Note that the results are slightly different for $T = 3$ and $T > 3$.

**Theorem 2** Suppose Assumptions 1 and 2 hold and $T \geq 3$. Then there is a unique solution to
the linear AB and quadratic AS moment restrictions except in the following three cases.

(i) Any value of $\alpha$ satisfies the linear AB and quadratic AS moment restrictions when $T \geq 3$
if and only if (7) holds and also

$$E(y_{it}^2) = E(y_{i0}^2), \quad t = 1, \ldots, T - 1.$$  \hfill (16)

(ii) There are two solutions to the linear AB and quadratic AS moment restrictions when
$T = 3$ if and only if (7) holds and

$$E(y_{i1}^2) \neq E(y_{i0}^2).$$ \hfill (17)

The solutions are the roots of the polynomial in (15b). If there is a double root, there is a
unique solution; otherwise there are two distinct solutions. One of the solutions is $\alpha = 1$ if and
only if (18) holds.

(iii) There are two distinct solutions to the linear AB and quadratic AS moment restrictions
when $T \geq 4$ if and only if (7) and (17) hold and also

$$E(y_{it-1} \Delta y_{it}) = 0$$ \hfill (18)

and

$$\exists \lambda \neq 1: \quad E(y_{it}^2) - E(y_{i,t-1}^2) = \lambda^{t-1}(E(y_{i1}^2) - E(y_{i0}^2)), \quad t = 2, \ldots, T - 1.$$ \hfill (19)
The two solutions are $\alpha = \lambda$ and $\alpha = 1$.

A proof can be found in Appendix A.2. Although Theorem 2 focuses mostly on identification failure, it is clear that the quadratic AS moment restrictions often provide unique identification of $\alpha_0$ when the linear AB moment restrictions do not. Furthermore, even if $\alpha_0$ is not uniquely identified, the quadratic AS moment restrictions often provide at least partial identification in the form of precisely two potential candidates. Hence, the quadratic AS moment restrictions can play an important role for identification as well as for efficient estimation.\footnote{It is clear from (15b) that (18) is a necessary condition for $\alpha = 1$ to be a solution to the moment restrictions. Moreover, (18) is implied by (16). Hence (18) is almost a necessary condition for identification failure. Only when $T = 3$ is it possible to have partial identification without (18). In this case $E(y_{iT-1}\Delta y_{iT})$ plays a role in determining the value of the “false” root.}

Regarding full unidentification, a stronger statement can be made. Because (7) and (16) are equivalent to $E((\Delta y_{it})^2) = 0$ for all $t = 1, \ldots, T - 1$, the conditions for full unidentification are satisfied if and only if the data are almost surely constant over time (up to $T - 1$) for all individuals. In terms of Figure 1, all elements in the matrix are identical except possibly the very last lower right element. It is not surprising that if there is no variation in the second moments of the data then it is not possible to separate the influence of $\alpha_0$, $y_{i0}$, and $c_i$. On the other hand, the converse means that if there is any variation in the second moments, then the data (and the minimal standard assumptions) provide some information about $\alpha_0$. When the quadratic AS moment restrictions are used, that information is either unique identification or two candidate values one of which is the true parameter.

Theorem 3 characterizes the DGPs for which the linear AB and the quadratic AS moment restrictions jointly do not identify $\alpha_0$. The motivation is partly to see precisely which DGPs are indistinguishable, and partly to see if some of the solutions described in Theorem 2 are incompatible with Assumptions 1 and 2 and can be ruled out.

\textbf{Theorem 3} Suppose Assumptions 1 and 2 hold and $T \geq 3$. Suppose (12) and (13) hold, so the linear AB moment restrictions fail to identify $\alpha_0$. Then the quadratic AS moment restrictions uniquely identify $\alpha_0$ except in the following cases.

The DGPs for which $\alpha_0$ is fully unidentified satisfy:
The DGPs for which $\alpha_0$ is partially identified satisfy either of:

(ii) $\alpha_0 = 1$, $\Psi = 0$, $E(v_{11}^2) \neq 0$, and $E(v_{1t}^2) = \lambda^{t-1}E(v_{11}^2)$ for all $t = 2, \ldots, T-1$ and some $0 \leq \lambda < 1$. The candidates are $\alpha = \alpha_0$ and $\alpha = \lambda$.

(iii) For $T = 3$: $0 \leq \alpha_0 < 1$, $\Psi \neq 0$, $E(v_{11}^2) = \alpha_0(1-\alpha_0)^{-1}\Psi$, and $E(v_{12}^2) = (\alpha_0^2 + \lambda - 1)(1-\alpha_0)^{-1}\Psi$ for some $0 \leq \lambda \leq 1$ and $\lambda \neq \alpha_0$. The candidates are $\alpha = \alpha_0$ and $\alpha = \lambda$.

For $T \geq 4$: $0 \leq \alpha_0 < 1$, $\Psi \neq 0$, $E(v_{11}^2) = \alpha_0(\alpha_0 - 1)^{-1}\Psi$, and $E(v_{1t}^2) = \alpha_0^{t-1}E(v_{11}^2)$ for $t = 2, \ldots, T-1$. The candidates are $\alpha = \alpha_0$ and $\alpha = 1$.

A proof of Theorem 3 can be found in Appendix A.3. The DGPs for which $\alpha_0$ is not identified is clearly a smaller set when the quadratic AS moment restrictions are used. For example, the linear AB moment restrictions do not involve $E(v_{1T-1}^2)$ and $E(v_{iT}^2)$, so these parameters do not play any role in determining when the linear AB moment restrictions fail to identify $\alpha_0$; however, Theorem 3 shows that the value of $E(v_{1T-1}^2)$ is important for identification through the quadratic AS moment restrictions.

Technically, it is interesting to note the duality between Cases (ii) and (iii) when $T \geq 4$, where the two candidates always belong to different cases. For example, the combination in (ii) with true value 1 and false candidate 0.5 is confused with the combination in (iii) with true value 0.5 and false candidate 1. For $T = 3$, the pattern is not quite as clear cut. Note also that

$$α_0^{T-1}\Psi + (α_0 - 1)E(v_{1T-1}^2) = 0$$

is a necessary condition for identification failure, except in Case (iii) when $T = 3$. Equation (20) is the $t = T - 1$ version of (13) in Theorem 1.

In case of partial identification, Theorem 3 shows that solutions outside the interval $[0, 1]$ can be ruled out. (This insight actually follows from the last part of Theorem 1.) For both $T = 3$ and $T \geq 4$, there is nothing in the moment restrictions that requires the “false”

---

If $T = 3$, the second root in Case (iii) simplifies to $α = 1 + E(v_{12}^2)/(E(v_{11}^2) - E(y_0c_1))$ if $α_0 = 0$, and to $α = (1-α_0^2) + α_0E(v_{12}^2)/E(v_{11}^2)$ if $0 < α_0 < 1$.\footnote{If $T = 3$, the second root in Case (iii) simplifies to $α = 1 + E(v_{12}^2)/(E(v_{11}^2) - E(y_0c_1))$ if $α_0 = 0$, and to $α = (1-α_0^2) + α_0E(v_{12}^2)/E(v_{11}^2)$ if $0 < α_0 < 1$.}
root to be between 0 and 1. That is, for some data generating processes there are solutions to the moment restrictions which are incompatible with the model specified in Assumption 1 and 2. In practice, such solutions must be discarded “manually”.

For $T \geq 4$, if it can be assumed that $|\alpha_0| < 1$ then identification is ensured also in cases where there are two distinct solutions to the moment restrictions. This follows because one of the two solutions must be $\alpha = 1$, which can be disregarded a priori. This leaves the other solution, $\alpha = \lambda$. Since $\alpha_0$ must be a solution, we have $\lambda = \alpha_0$ and $|\lambda| < 1$.

Table 1 provides an overview of the conclusions of Theorems 1 and 3. To emphasize completeness, the presentation is different. The DGPs are grouped by whether or not $E(v_{it}^2) = 0$ for all $t = 1, \ldots, T-2$, whether or not $\Psi = 0$, and whether $\alpha_0 = 0$, $\alpha_0 = 1$, or $\alpha_0 \neq 0, \alpha_0 \neq 1$. For each case, the table shows conditions under which the linear second moment restrictions fail to identify $\alpha_0$ and information provided by the quadratic second moment restrictions.

The results can be summarized from a practical perspective as follows. If there is a unique solution to the moment restrictions, then that solution is $\alpha_0$. In this case identification may come from either the linear AB or the quadratic AS moment restrictions or both. If there are more than two solutions to the moment restrictions, then it must be the case that $\Psi = 0$ and $E(v_{it}^2) = 0$ for all $t = 1, \ldots, T-2$. If there are exactly two solutions to the moment restrictions, the linear AB moment restrictions must have failed to identify $\alpha_0$, but the quadratic AS moment restrictions provide partial identification.

Although estimation is not of our immediate interest in this paper, notice that in practice it may be difficult to determine whether $\alpha_0$ is uniquely or partially identified. Since the moment restrictions are quadratic, the GMM objective function is a quartic function. This means that there are either one or two real local minima, and that there could be two local minima even if $\alpha_0$ is uniquely identified. (The probability of a single local minimum vanishes asymptotically if $\alpha_0$ is partially identified.) That is, the finding of multiple local minima in a particular application is not sufficient to conclude that $\alpha_0$ is not uniquely identified.
5 First moment restrictions

As mentioned, Ahn and Schmidt (1995) showed that the linear AB and quadratic AS moment restrictions constitute the complete set of available restrictions on the second moments of the data implied by Assumptions 1 and 2. However, it is obvious that there are also first moment restrictions available under Assumption 1. Until very recently the first moment restrictions have essentially been ignored in the literature. This is puzzling because their potential contribution to both identification and efficiency is great. As far as we are aware Han and Kim (2014) and Ahn and Kitazawa (2014) are the only authors to explicitly consider them. They focused on estimation and efficiency and did not discuss identification.

Ignoring the first moments creates a certain lack of invariance in the identification results presented so far. For example, suppose $E(y_{it} \Delta y_{it+1}) = 0$ for some $t$. Then the lack of invariance manifests itself in the fact that this does not imply $E(y^*_it \Delta y^*_it+1) = 0$, where $y^*_it$ is $y_{it}$ plus some constant. This is rather awkward, because it suggests that some identification problems can be fixed simply by adding a constant to the data. Clearly it is not desirable that identification depends on an arbitrary normalization of the data.

The lack of invariance (or the fact that adding a constant helps) is an indication that there is information in the first moments that should be utilized. Invariance, and in some cases identification, is ensured by using first moment restrictions together with the second moment restrictions. We discuss extended models which include time effects below. In the simple model with no time effects, the first moment restrictions are linear and can be written as

$$E(\Delta y_{it}) - E(\Delta y_{it-1})\alpha = 0, \quad t = 2, \ldots, T. \quad (21)$$

By Assumption 1, (21) is satisfied for $\alpha = \alpha_0$. We have written them here as $T-1$ equations in first differences. They could be written equivalently as $T$ equations in levels, but then involve an additional parameter, $E(c_i)$, which acts as a common constant term across the time periods.

By (21) and Assumption 1, a necessary and sufficient condition for identification failure
is that

\[ E(\Delta y_{11}) = 0. \]  \hfill (22)

The data are said to be mean stationary if \( E(y_{it}) = E(y_{i0}) \) for all \( t = 1, \ldots, T \). Assumption 1 implies that the data are mean stationary if and only if (22) holds.

Theorem 4 characterizes the DGPs for which the first moment restrictions do not identify \( \alpha_0 \).

**Theorem 4** Suppose Assumption 1 holds. If \( T \geq 2 \), then a necessary and sufficient condition for the first moment restrictions to fail to identify \( \alpha_0 \) is that

\[ (\alpha_0 - 1)E(y_{i0}) + E(c_i) = 0. \]  \hfill (23)

A proof, which is trivial, is given in Appendix A.4. Note that if \( \alpha_0 = 1 \), (23) implies \( E(c_i) = 0 \), which is consistent with but weaker than \( \Psi = E(c_i^2) = 0 \) in Theorem 3.

Often an empirical application requires constant terms which change over time. The most flexible model in this regard includes separate time-specific constant terms, \( \tau_1, \ldots, \tau_T \), for each time period (see e.g. Holtz-Eakin et al., 1988). To analyze such a model, Assumption 1 is replaced with Assumption 3.

**Assumption 3** The random variables \( \{(y_{i0}, \ldots, y_{iT}) : i = 1, \ldots, n\} \) are independent across individuals, have finite means, and satisfy

\[ y_{it} = y_{it-1} + \alpha_0 + \tau_t + c_i + v_{it}, \quad t = 1, \ldots, T, \]  \hfill (24)

with \( E(v_{it}) = 0 \) for \( t = 1, \ldots, T \).

It is easy to see that the \( E(c_i) \) and \( \tau_1, \ldots, \tau_T \) are not separately identified, but that is not a concern here where we focus on \( \alpha_0 \).

However, Assumption 3 adds \( T \) parameters, and they absorb all information in the \( T \) first moment restrictions. To see this, define the mean-adjusted variables \( \tilde{y}_{it} = y_{it} - E(y_{it}) \)
for $t = 0, \ldots, T$. If the original variables satisfy Assumption 3, then the mean-adjusted variables satisfy Assumption 1. The complete set of moment restrictions is therefore given by (5) and (6) substituting $y_{it}$ with $\bar{y}_{it}$ for all $t$ while (21) is replaced with

$$E(\Delta y_{it}) - E(\Delta y_{it-1}) \alpha - \Delta \tau_i = 0, \quad t = 2, \ldots, T.$$ (25)

The time effects appear only in (25), which constitute $T - 1$ equations with $T$ unknowns. Therefore, if (5) and (6) identify $\alpha_0$, then (25) exactly identifies $\Delta \tau_2, \ldots, \Delta \tau_T$. The first moment restrictions therefore cannot provide exploitable information about $\alpha_0$.

To sum up, if the model does not have time effects and Assumptions 1 and 2 hold, then the first moments may provide strong identifying information. Specifically, the first moments are informative if the data are not mean stationary. Furthermore, including the first moments solves the lack of invariance problem in the linear AB and quadratic AS moment restrictions. If the model has separate time effects for each time period and Assumptions 3 and 2 hold, then the first moment restrictions identify the time effects (up to a common constant).

For completeness we mention that there is a range of models between specifications (1) and (24) which have less flexible time effects. For example, consider a model with a linear time trend, say $y_{it} = y_{it-1} \alpha_0 + \gamma t + c_i + v_{it}$ for $t = 1, \ldots, T$. Here the first and second moment restrictions may all contribute to identifying $\alpha_0$, $\gamma$, and $E(c_i)$.

### 6 Supplementary assumptions

Two supplementary assumptions have been considered in the literature (see Ahn and Schmidt, 1995), namely “homoskedasticity” in the sense that

$$E(v_{it}^2) = E(v_{i1}^2), \quad t = 1, \ldots, T,$$ (26)

7Following the majority of the literature, this paper utilizes the second moments in their raw rather than central form. If the second central moments are used instead, no adjustment for time effects is needed.
and “stationarity” in the sense that

\[ E(y_{it}c_i) = E(y_{i0}c_i), \quad t = 1, \ldots, T. \]  

(27)

It is not difficult to show that stationarity holds if and only if \( \Psi = 0 \) (see Appendix B). These assumptions eliminate many DGPs a priori, including some where \( \alpha_0 \) is not identified, as well as provide additional moment restrictions that can be used for identification and estimation. For simplicity, we focus on the case where \( T \geq 4 \) and \( E(v_i^2) > 0 \) in this section.

Under homoskedasticity, the linear AB moment restrictions fail to identify \( \alpha_0 \) if and only if \( \alpha_0 = 1 \) and \( E(c_i^2) = 0 \). This follows because \( \alpha_0 = 1 \) and \( \Psi = 0 \) is the only solution to (13) in Theorem 1, and this in turn implies \( E(c_i^2) = 0. \)\(^8\) Other cases of nonidentification are not compatible with homoskedasticity and hence, by assumption, cannot occur. Thus, identification is ensured in Alvarez and Arellano’s (2003) theoretical study, since they assume homoskedasticity, stationarity, and \( |\alpha_0| < 1 \). Theorem 3 shows that if both the linear AB and the quadratic AS moment restrictions are used, then \( \alpha_0 \) is identified without the assumption \( |\alpha_0| < 1 \). This follows because the cases with partial identification cannot occur under homoskedasticity. In practice, few empirical studies assume homoskedasticity.

The results under stationarity are quite similar. The linear AB moment restrictions fail to identify \( \alpha_0 \) if and only if \( \alpha_0 = 1 \) and \( E(c_i^2) = 0 \). This follows because \( \alpha_0 = 1 \) is the only solution to (13) when \( \Psi = 0 \), and because \( \alpha_0 = 1 \) and \( \Psi = 0 \) imply \( E(c_i^2) = 0 \). Again nonidentification happens only in the random walk case, as other cases are ruled out by assumption. Theorem 3 shows that if both the linear AB and the quadratic AS moment restrictions are used, then \( \alpha_0 \) is uniquely identified without assuming \( |\alpha_0| < 1 \). This follows because the cases with partial identification are cannot occur under stationarity.

While homoskedasticity is rarely imposed in practice, many empirical studies assume stationarity. Stationarity implies additional linear “level” moment restrictions that are used in the popular “system GMM” estimator (see e.g. Arellano and Bover, 1995; Blundell and

---

\(^8\)Since it is the only possibility (assuming \( T \geq 4 \) and \( E(v_i^2) > 0 \)), the finding that \( \alpha_0 \) is “unidentified” implies that \( \alpha_0 = 1 \), so strictly speaking \( \alpha_0 \) is identifiable from the linear AB moment restrictions.
\[ E((y_{it} - \alpha y_{it-1}) \Delta y_{it-1}) = E(y_{it} \Delta y_{it-1}) - \alpha E(y_{it-1} \Delta y_{it-1}) = 0, \quad t = 2, \ldots, T. \] (28)

Since \( \alpha_0 \) is already uniquely identified under stationarity by the linear AB and quadratic AS moment restrictions, the level moment restrictions cannot contribute additionally to identification. (However, the level moment restrictions can contribute to efficiency.)

Interestingly, if the stationarity assumption is not satisfied and the linear AB moment restrictions do not identify \( \alpha_0 \), then we must have \( \alpha_0 \neq 1 \) by Theorem 1; however, it can be shown that either \( \alpha = 1 \) is the only solution to the level moment restrictions, or the level moment restrictions do not have a common solution. In practice, this means that if the stationarity assumption does not hold, then the system GMM estimator is likely to be biased towards 1 or a test of overidentification restrictions will be rejected. Recently authors have argued that the stationarity assumption may not be appropriate (see e.g. Hayakawa, 2009; Calzolari and Magazzini, 2013).

A general treatment of stationarity and level moment restrictions is provided in Appendix B.

7 Simulation results

In this section, we present simulation results which illustrate the potential consequences of identification failure for GMM estimation. The points are simple. First, if the linear AB moment restrictions do not identify \( \alpha_0 \), but the quadratic AS moment restrictions do, then the difference GMM estimator has poor properties, while the nonlinear GMM estimator based on the first, linear AB, and the quadratic AS moment restrictions performs much better. Second, the poor properties of a GMM estimator extends to a neighborhood around points of nonidentification. Third, there are cases where the difference GMM estimator is inconsistent which can be solved simply by adding a constant to all observed data.

\(^9\)Identification is also ensured by using the linear AB moment restrictions together with the level moment restrictions.
7.1 DGP 1: $\alpha_0 = 0.5$

Table 2 presents the results of three small Monte Carlo experiments with $\alpha_0 = 0.5$. There are 5000 Monte Carlo replications in each experiment, and all the GMM estimators are based on the optimal weight matrix. Further details about the DGPs are given in the table notes. Note that the parameters are chosen to be ordinary-looking.

The first panel of Table 2 shows the properties of GMM estimators based on different combinations of the first and second moment restrictions. For the first and second data generating processes (DGP 1.1 and DGP 1.2), neither the first moment restrictions nor the linear AB moment restrictions identify $\alpha_0$. The difference GMM estimator is therefore inconsistent. In contrast, the quadratic AS moment restrictions uniquely identify $\alpha_0$, and the performance of the nonlinear GMM estimator based on the linear AB and quadratic AS moment restrictions is much better, as shown in the second line of results. The remaining results in the first panel of Table 2 simply confirm that adding a constant to the observed data or adding the first moment restrictions makes little difference when the data generating process is mean stationary. The middle panel in Table 2 shows results for the same data generating process when the sample size is larger. As expected, the performance of consistent estimators improve, while that of inconsistent estimators remain poor. In sum, the results in the top and middle panels in Table 2 illustrate the potential benefit of exploiting the quadratic AS moment restrictions for identification and estimation.

Poor performance is important not only at points of nonidentification, but also in a neighborhood around such points. To illustrate this, Figure 2 shows the RMSE for four estimators as $\alpha_0$ varies near 0.5, when all other parameters are as in DGP 1.2. For the GMM estimators which use only the first and the linear AB moment restrictions, the RMSE is clearly higher the closer $\alpha_0$ is to 0.5. On the other hand, when the quadratic AS moment restrictions are used, there is no problem.

The third panel of Table 2 show results for a data generating process (DGP 1.3) which is not mean stationary, so the first moment restrictions identify $\alpha_0$. Again, the parameters are chosen such that the linear AB moment restrictions do not identify $\alpha_0$, but the quadratic AS moment restrictions uniquely identify $\alpha_0$. The performance of the standard difference and
nonlinear GMM estimators are quite similar to the previous experiment. However, the third line shows that performance of the standard difference GMM estimator improves dramatically if the data are translated by adding a constant, 5 in this instance, to all the observed values. This experiment therefore illustrates how a data analyst who relies on the difference GMM estimator can shoot himself in the foot by ignoring the first moment restrictions. The fourth and fifth lines show that explicitly exploiting the first moment restrictions is even better than adding 5. This is because the number of informative moment restrictions is larger. Finally, increasing the sample size to \( n = 5000 \) improves the performance of consistent estimators, while the properties of inconsistent estimators remain poor (not shown in the table). To sum up, the results in the third panel in Table 2 illustrate the potential problem if the first moment restrictions are ignored as well as further emphasize the usefulness of the quadratic moment restrictions.

### 7.2 DGP 2: \( \alpha_0 = 1 \)

To show how the quadratic AS moment restrictions help identifying \( \alpha_0 \) when \( \alpha_0 = 1 \), we present simulation results for three experiments in Table 3. Again there are 5000 Monte Carlo replications in each experiment, all the GMM estimators are based on the optimal weight matrix, and details about the data generating processes are given in the table notes. In all three experiments, the linear AB moment restrictions fail to identify \( \alpha_0 \), while the quadratic AS moment restrictions provide identification.

The two GMM estimators based only on the first and the linear AB moment restrictions are inconsistent in all experiments, so it is not surprising that their properties are rather poor. (Also adding a constant to all observed data does not help in these experiments.) In contrast, the nonlinear GMM estimator performs much better, as expected. In the first (and second) panel, the data generating process is such that the \( v_{it} \) are homoskedastic. It is known that the GMM estimators converge very slowly in this particular case (see e.g. Bun and Kleibergen, 2013). This evidenced in the second and third panels, which show results when the sample size is increased to \( n = 5000 \) and when \( v_{it} \) are heteroskedastic, respectively. The convergence rate is root-\( n \) for the latter design. The numbers in the second and third
panel are more similar to each other than to those in the first panel.

Figure 3 shows that the poor performance of both the standard difference GMM estimator and the estimator based on the first and linear AB moment restrictions extends to a neighborhood around the point of unidentification. As in Figure 2, the RMSE for these estimators is the higher the closer $\alpha_0$ is to 1, while the GMM estimators which exploit the quadratic AS moment restrictions are fine.

7.3 DGP 3: on the edge

According to Theorem 1, the linear AB moment restrictions fail to identify $\alpha_0$ in some cases when $E(v_{it}^2) = 0$ for all $t = 1, \ldots, T - 2$. Theorem 3 implies that the quadratic AS moment restrictions identify $\alpha_0$ in these situations, provided $E(v_{i(t-1)}^2) \neq 0$. We briefly show that the estimation problem extends to a neighborhood near points of identification failure.

We consider the following data generating processes: $T = 4$ and $y_{i0}, c_i, v_{i1}, \ldots, v_{iT}$ normally distributed; $\alpha_0 = 0.5$, $E(y_{i0}) = 0$, $E(c_i) = 0$, $E(y_{i0}^2) = 4$, $E(c_i^2) = 1$, $E(y_{i0}c_i) = 2$, $E(v_{i1}^2) = \rho^2$, $E(v_{i2}^2) = \rho^2$, $E(v_{i3}^2) = 1$, and $E(v_{i4}^2) = 1$, where $\rho$ varies across experiments. The simulations have 5000 Monte Carlo replications and use the optimal weight matrix. Figure 4 shows the RMSE for the four GMM estimators when $\rho$ varies from 0 to 1. The poor performance of the difference GMM estimator (even if supplemented with first-moment restrictions) is obvious. The GMM estimators that use the quadratic moment restrictions perform well.

8 Concluding remarks

In this paper we revisited the familiar AR(1) dynamic panel data model with fixed effects. We showed that the linear AB moment restrictions available under standard minimal assumptions do not guarantee identification of $\alpha_0$, even if it can be assumed that $|\alpha_0| < 1$. We demonstrated that the quadratic AS moment restrictions and the first moment restrictions can provide identification when the linear AB moment restrictions fail. Specifically, the quadratic AS moment restrictions uniquely identify $\alpha_0$ in many cases and, if the data
exhibits any variation across time at all, they always provide at least partial identification in the form of two candidate values. The first moment restrictions, which hitherto have been virtually ignored in the literature, provide important identification when the data are not mean stationary, and in particular eliminate embarrassing problems due to inappropriate data normalization. The paper presented complete identification results both in term of the moments of the observed data and in terms of the underlying DGP.

Since the popular difference GMM estimator is inconsistent when the linear AB moment restrictions do not provide identification, our findings have important implications for empirical work. Our simulation results showed that the properties of various GMM estimators can be very poor not only when the moment restrictions used for estimation do not identify \( \alpha_0 \), but also for nearby DGP. The simulations also evidenced the potential benefits of using the quadratic AS moment restrictions in addition to the linear AB moment restrictions.

The focus on the simple AR(1) model without covariates has enabled us to clarify the identification issues without having to deal with too many special cases. Practical applications usually warrant more complicated models. Useful extensions of this paper therefore include examining models with higher-order dynamics and models with covariates.

**A Appendix**

The following relationships are used in several of the proof below. Repeated substitution in (1) yields

\[
y_{it} = y_{i0} \alpha_t^0 + c_i \sum_{j=1}^{t} \alpha_{0}^{t-j} + \sum_{j=1}^{t} \nu_{ij} \alpha_{0}^{t-j}, \quad t = 1, \ldots, T.
\]  

(29)

Also from (1) we have

\[
\Delta y_{it} = \Delta y_{i1} \alpha_0^{t-1} + \sum_{j=0}^{t-2} \Delta \nu_{it-j} \alpha_0^j, \quad t = 2, \ldots, T,
\]

(30)
and from (29) we have

\[
\Delta y_{it} = \begin{cases} 
  y_{i0}(\alpha_0 - 1) + c_i + v_{it}, & \text{if } t = 1, \\
  y_{i0}\alpha_0^{t-1}(\alpha_0 - 1) + c_i\alpha_0^{t-1} + \sum_{j=1}^{t-1} v_{ij}\alpha_0^{t-1-j}(\alpha_0 - 1) + v_{it}, & \text{if } t = 2, \ldots, T.
\end{cases} 
\]  

\[ (31) \]

\[ \text{\textit{A.1 Proof of Theorem 1}} \]

Define \( \Pi = (\alpha_0 - 1)E(y_{i0}^2) + E(y_{i0}c_i) \).

**Lemma 1** Suppose Assumptions 1 and 2 hold. Then \( \Pi = 0 \) implies \( \Psi \geq 0 \).

**Proof of Lemma 1** If \( \Pi = 0 \), then \((1-\alpha_0)^2E(y_{i0}^2)^2 = E(y_{i0}c_i)^2 \), and \( E(y_{i0}c_i)^2 \leq E(y_{i0}^2)E(c_i^2) \) by the Cauchy-Schwarz inequality. Also \( \Pi = 0 \) implies \( \Psi = -(\alpha_0 - 1)^2E(y_{i0}^2) + E(c_i^2) \), so if \( E(y_{i0}^2) > 0 \) then the previous inequality yields \( \Psi \geq 0 \). If \( E(y_{i0}^2) = 0 \) then \( \Psi = E(c_i^2) \geq 0 \). The conclusion follows. \( \blacksquare \)

To prove Theorem 1, note that by (30), (31), and Assumption 2 we have

\[
E(y_{is}\Delta y_{is+1}) = \begin{cases} 
  E(y_{i0}^2)(\alpha_0 - 1) + E(y_{i0}c_i), & \text{if } s = 0, \\
  \alpha_0^sE(y_{is-1}\Delta y_{is}) + (\alpha_0 - 1)E(v_{is}^2) \\
  + \alpha_0^s[(\alpha_0 - 1)E(y_{i0}c_i) + E(c_i^2)] & \text{if } s = 1, \ldots, T - 2.
\end{cases} 
\]  

\[ (32) \]

It follows that the linear AB moment restrictions fail to identify \( \alpha_0 \) if and only if

\[
(\alpha_0 - 1)E(y_{i0}^2) + E(y_{i0}c_i) = 0
\]  

\[ (33) \]

and

\[
\alpha_0^t[(\alpha_0 - 1)E(y_{i0}c_i) + E(c_i^2)] + (\alpha_0 - 1)E(v_{it}^2) = 0, \quad t = 1, \ldots, T - 2,
\]  

\[ (34) \]

where the latter is empty if \( T = 2 \).

It is obvious that any value of \( \alpha \) is compatible with \( \Pi = 0 \) when \( T = 2 \).
To prove that identification failure is possible for any $\alpha_0$ when $\Psi = 0$ and $E(y_{it}^2) = 0$ for all $t = 1, \ldots, T - 2$ when $T \geq 3$, it is sufficient to show that the combination $\Pi = 0$ and $\Psi = 0$ is compatible with any value of $\alpha_0$. Let $w_i$ be standard normal random variable. If $\alpha_0 \neq 1$, define $y_{i0} = -w_i(\alpha_0 - 1)^{-1}$ and $c_i = w_i$. Then $E(y_{i0}^2) = (\alpha_0 - 1)^{-2}$, $E(c_i^2) = 1$, and $E(y_{i0}c_i) = -(\alpha_0 - 1)^{-1}$. If $\alpha_0 = 1$, define $y_{i0} = w_i$ and $c_i = 0$. Then $E(y_{i0}^2) = 1$, $E(c_i^2) = 0$ and $E(y_{i0}c_i) = 0$. In either case, it is easy to see that $\Pi = (\alpha_0 - 1)E(y_{i0}^2) + E(y_{i0}c_i) = 0$ and $\Psi = (\alpha_0 - 1)E(c_iy_{i0}) + E(c_i^2) = 0$.

To prove identification failure implies $0 \leq \alpha_0 \leq 1$ when $T \geq 3$ and either $\Psi \neq 0$ or $E(v_{it}^2) \neq 0$ for some $t = 1, \ldots, T - 2$, note that (13) can be written $\alpha_0^2 = -(\alpha_0 - 1)E(v_{it}^2)$ for all $t = 1, \ldots, T - 2$. By Lemma 1, $\Psi < 0$ is not possible when $\alpha_0$ is not identified. If $\Psi > 0$ and $E(v_{it}^2) = 0$ for all $t = 1, \ldots, T - 2$, then we must have $\alpha_0 = 0$. If $\Psi = 0$ and $E(v_{it}^2) > 0$ for some $t = 1, \ldots, T - 2$, then we must have $\alpha_0 = 1$. If $\Psi > 0$ and $E(v_{it}^2) > 0$ for some $t = 1, \ldots, T - 2$, note that $\alpha_0 = 0$ and $\alpha_0 = 1$ are not possible, so we must have $E(v_{it}^2) = -\alpha_0^2(\alpha_0 - 1)^{-1}\Psi > 0$ for $t = 1, \ldots, T - 2$, and the two sides of the equation have opposite signs unless $0 < \alpha_0 < 1$.

To prove that identification failure is possible for any value $0 \leq \alpha_0 \leq 1$, see the example in the text in Section 3.

A.2 Proof of Theorem 2

To prove Theorem 2, define $C_{it} = E(y_{it}^2) - E(y_{it-1}^2)$ for $t = 1, \ldots, T$. Then (15) can be written

$$C_{it} - (C_{it} + C_{it-1})\alpha + C_{it-1}\alpha^2 = 0, \quad t = 2, \ldots, T - 2,$$

$$E(y_{iT-1}\Delta y_{iT}) + C_{iT-1} - (C_{iT-1} + C_{iT-2})\alpha + C_{iT-2}\alpha^2 = 0. \quad (35a)$$

Consider first (35b). Given that there must be at least one solution (i.e. $\alpha = \alpha_0$), there are three possible scenarios: if $C_{iT-1} \neq 0$ and $C_{iT-2} = 0$, then (35b) has a unique solution; if $C_{iT-1} = 0$ and $C_{iT-2} = 0$, then (35b) implies $E(y_{iT-1}\Delta y_{iT}) = 0$, so any $\alpha$ solves (35b); and if $C_{iT-2} \neq 0$, then either (35b) has a unique solution (double root) or (35b) has two distinct solutions. Note that if $C_{iT-2} \neq 0$, then (35b) has solution $\alpha = 1$ if and only if
\( \mathbb{E}(y_{iT-1} \Delta y_{iT}) = 0. \)

For \( T = 3 \), (35b) is the only quadratic restriction and the conclusion of the theorem follows from the above arguments. In particular, (16) corresponds to \( C_{iT-1} = 0 \) and \( C_{iT-2} = 0 \) and (17) corresponds to \( C_{iT-2} \neq 0 \).

For \( T \geq 4 \), we need to consider all equations in (35a) and (35b) together. To begin, consider a representative equation in (35a): if \( C_{it} \neq 0 \) and \( C_{it-1} = 0 \), then \( \alpha = 1 \) is a unique solution; if \( C_{it} = 0 \) and \( C_{it-1} = 0 \) then any \( \alpha \) is a solution; and if \( C_{it-1} \neq 0 \), then there are two solutions, \( \alpha = 1 \) and \( \alpha = C_{it}/C_{it-1} \).

Therefore, \( \alpha_0 \) is fully unidentified if and only if all equations (35a) and (35b) are satisfied by any \( \alpha \); that is, if and only if \( C_{it} = 0 \) for all \( t = 2, \ldots, T-1 \); that is,

\[
\mathbb{E}(y_{it}^2) = \mathbb{E}(y_{i0}^2), \quad t = 1, \ldots, T-1. \quad (36)
\]

Note that \( \mathbb{E}(y_{iT-1} \Delta y_{iT}) = 0 \) is implied by (35b) in this case.

Furthermore, \( \alpha_0 \) is partially identified if and only if at least one of the equations in (35a) and (35b) have exactly two distinct solutions and these two solutions satisfy all equations. Possibly some equations may be satisfied by any value of \( \alpha \).

Suppose \( C_{it-1} = 0 \) and \( C_{it} \neq 0 \) for some \( t = 2, \ldots, T-1 \). Then the equation for that \( t \) uniquely identifies \( \alpha_0 \), and this pattern is thus incompatible with partial identification. Therefore, necessary conditions for partial identification are that \( C_{is} = 0 \) whenever \( C_{it-1} = 0 \) for all \( s = t, \ldots, T-1 \) and that \( C_{is} \neq 0 \) whenever \( C_{it} \neq 0 \) for all \( s = 1, \ldots, t-1 \). Since partial identification requires \( C_{it-1} \neq 0 \) for some \( t = 2, \ldots, T-1 \), it follows that \( C_{i1} \neq 0 \) is a necessary condition. Define \( \lambda = C_{i2}/C_{i1} \). Then a necessary condition for partial identification is that the equation for \( t = 2 \) in (35a) has solutions \( \alpha = 1 \) and \( \alpha = \lambda \), with \( \lambda \neq 1 \).

Suppose \( \lambda \neq 0 \). Then \( C_{i2} \neq 0 \), which implies the equation for \( t = 3 \) also has exactly two solutions. These solutions are \( \alpha = 1 \) and \( \alpha = \lambda \) if and only if \( C_{i3} = \lambda C_{i2} \). (Recall that if \( C_{iT-2} \neq 0 \), then (35b) has solution \( \alpha = 1 \) if and only if \( \mathbb{E}(y_{iT-1} \Delta y_{iT}) = 0 \), in which case (35b) has the same form as (35a).) So \( C_{i2} \neq 0 \) is a necessary condition for partial identification with \( \lambda \neq 0 \). By iterating this argument, partial identification arises when \( C_{it} = \lambda C_{it-1} \) for all \( t = 2, \ldots, T-1 \).
Suppose \( \lambda = 0 \). Then \( C_{i2} = 0 \), which implies \( C_{it} = 0 \) for all \( t = 2, \ldots, T-1 \). It follows that the equations for \( t = 3, \ldots, T - 1 \) must be satisfied for any value of \( \alpha \). Note that \( C_{iT-2} = 0 \) and \( C_{iT-1} = 0 \) implies \( E(y_{it-1}\Delta y_{it}) = 0 \) by (35b).

Putting together the arguments for \( \lambda \neq 0 \) and \( \lambda = 0 \), partial identification occurs if and only if all equations have exactly the same two distinct solutions, namely \( \alpha = \lambda \) and \( \alpha = 1 \) with \( \lambda \neq 1 \), and this requires \( C_{i1} \neq 0 \), \( C_{it} = \lambda C_{it-1} \) for all \( t = 2, \ldots, T - 1 \) and \( E(y_{iT-1}\Delta y_{iT}) = 0 \).

### A.3 Proof of Theorem 3

To characterize the DGPs for which identification of \( \alpha_0 \) fails and prove Theorem 3, we begin by partitioning the DGPs into three groups. For each group, we then express the quadratic moment restrictions (14) in terms of the parameters of the data generating process, and consider the roots of the resulting equation. Throughout we consider only the case where the linear AB moment restrictions do not identify \( \alpha_0 \) (Theorem 1).

**Case \( \Psi = 0 \) and \( E(v^2_{it}) = 0 \) for \( t = 1, \ldots, T-2 \)**

First, \( E(v^2_{it}) = 0 \) implies \( v_{it} = 0 \) and \( \Delta y_{it} = \alpha_0^{t-1}\Delta y_{i1} \) for all \( t = 1, \ldots, T-2 \). Second, \( \Psi = 0 \) implies \( E(c_i\Delta y_{i1}) = 0 \). Third, the fact that the linear AB moment restrictions fail to identify \( \alpha_0 \) implies \( E(y_{i0}\Delta y_{i1}) = 0 \). These results plus the fact that \( y_{it} \) is a linear combination of \( c_i \) and \( y_{i0} \) then imply

\[
E(y_{is}\Delta y_{it}) = \alpha_0^{s-1}E(y_{i0}\Delta y_{i1}) = 0, \quad s = T-1, T, \quad t = 1, \ldots, T-2.
\]

(37)

It follows that \( E(y_{iT}\Delta y_{it}) = E(y_{iT-1}\Delta y_{it}) = E(y_{iT}\Delta y_{iT-1}) = E(y_{iT-1}\Delta y_{iT-1}) = 0 \), so the quadratic AS moment restrictions for \( t = 2, \ldots, T-2 \) in (14) are satisfied for any value of \( \alpha \). For \( t = T-1 \) we have

\[
E(y_{iT}\Delta y_{iT-1}) = E[(\alpha_0^2y_{iT-2} + (\alpha_0 + 1)c_i + v_{iT} + \alpha_0v_{iT-1})(\alpha_0^{T-2}\Delta y_{i1} + v_{iT-1})] = \alpha_0 E(v^2_{iT-1}).
\]

(38)
Similarly, \( E(y_{iT-1} \Delta y_{iT-1}) = E(v^2_{iT-1}) \) and \( E(y_{iT} \Delta y_{iT-2}) = E(y_{iT-1} \Delta y_{iT-2}) = 0 \). The corresponding quadratic AS moment restriction becomes

\[
\alpha_0 E(v^2_{iT-1}) - E(v^2_{iT-1}) \alpha = 0. \tag{39}
\]

Therefore, the quadratic AS moment restrictions uniquely identify \( \alpha_0 \) if \( E(v^2_{iT-1}) \neq 0 \); otherwise they provide no information about \( \alpha_0 \).

**Case \( \Psi = 0 \) and \( E(v^2_{iT}) \neq 0 \) for some \( t = 1, \ldots, T - 2 \)**

From (13) with \( \Psi = 0 \), we get \( \alpha_0 = 1 \). By definition of \( \Psi \), we then get \( E(c_i^2) = 0 \), and consequently \( c_i = 0 \). It follows that \( \alpha_0 = 1 \), we have that \( y_{it} = y_{i0} + v_{i1} + \cdots + v_{it} \) and \( \Delta y_{it} = v_{it} \) for \( t = 2, \ldots, T - 1 \). The quadratic moment restrictions in (14) become

\[
E(v^2_{it}) - (E(v^2_{it}) + E(v^2_{it-1})) \alpha + E(v^2_{it-1}) \alpha^2 = 0, \quad t = 2, \ldots, T - 1. \tag{40}
\]

If \( T = 3 \), there are roots at \( \alpha = 1 \) and \( \alpha = E(v^2_{i2})/E(v^2_{i1}) \), so \( \alpha_0 \) is uniquely identified if and only if \( E(v^2_{i2}) = E(v^2_{i1}) \) and partially identified otherwise. (Note that \( E(v^2_{i1}) > 0 \) by assumption.) If \( T \geq 4 \), there exists \( \sigma \) \( (1 \leq \sigma \leq T - 2) \) such that \( E(v^2_{i\sigma}) > 0 \). It follows that the moment restriction for \( t = \sigma + 1 \) has roots at 1 and \( E(v^2_{i\sigma+1})/E(v^2_{i\sigma}) \). It follows that \( \alpha_0 \) is at least partially identified. Moreover, the full set of quadratic moment restrictions have two distinct common roots if and only if there is \( \lambda \neq 1 \) such that \( \lambda E(v^2_{iT-1}) = E(v^2_{iT}) \) for all \( t = 2, \ldots, T - 1 \). It follows that \( \alpha_0 \) is uniquely identified unless such \( \lambda \neq 1 \) exists; otherwise there are roots at \( \alpha = 1 \) and \( \alpha = \lambda \).

**Case \( \Psi \neq 0 \)**

First, when \( \Psi \neq 0 \) and the linear AB moment restrictions fail to identify \( \alpha_0 \), then (13) implies \( \alpha_0 \neq 1 \), and therefore Theorem 1 implies that \( 0 \leq \alpha_0 < 1 \). By (13), we then have \( E(v^2_{it}) = \alpha_0^t(1 - \alpha_0)^{-1} \Psi \) for \( t = 1, \ldots, T - 2 \), while \( E(v^2_{iT-1}) \) and \( E(v^2_{iT}) \) are unrestricted. Second, by (30) and Assumption 2, we also have that \( E(c_i \Delta y_{it}) = \alpha_0^{t-1} \Psi \) for \( t = 1, \ldots, T \).
Third, for any \( t = 2, \ldots, T - 1 \) repeated substitution gives

\[
y_{iT-1} = \alpha_0^{T-t+1} y_{it-2} + c_i \sum_{j=1}^{T-t+1} \alpha^{j-1} + \sum_{j=t-1}^{T-1} v_j \alpha^{T-j-1}. \tag{41}
\]

Fourth, by (41) and the first intermediate result, we have

\[
E(y_{iT-1} \Delta v_{it}) = \alpha_0^{T-t-1} E(v_{it}^2) - \alpha_0^{T-t} E(v_{it-1}^2) = \alpha_0^{T-t-1} \Upsilon_t, \quad t = 2, \ldots, T - 1, \tag{42}
\]

where \( \Upsilon_t = E(v_{it}^2) - \alpha_0'(1 - \alpha_0)^{-1} \psi \) for \( t = 1, \ldots, T \) is defined as the deviation of \( E(v_{it}^2) \) from the geometric trend. Note that \( \Upsilon_t = 0 \) for \( t = 1, \ldots, T - 2 \), while \( \Upsilon_{T-1} \) and \( \Upsilon_T \) are unrestricted.

By (41), by noting that \( E(y_{it-2} \Delta y_{it-1}) = 0 \) for all \( t = 2, \ldots, T - 1 \) since the linear moment restrictions do not identify \( \alpha_0 \), and by using the first and the second results, we have

\[
E(y_{iT-1} \Delta y_{it-1}) = \alpha_0^{T-t-2}(1 - \alpha_0)^{-1} \psi, \quad t = 2, \ldots, T - 1. \tag{43}
\]

Similar manipulations yield

\[
E(y_{iT-1} \Delta y_{it}) = \alpha_0^{T-t-1}(1 - \alpha_0)^{-1} \psi + \alpha_0^{T-t-1} \Upsilon_t, \quad t = 2, \ldots, T - 1, \tag{44}
\]

\[
E(y_{iT} \Delta y_{it-1}) = \alpha_0^{T-t-2}(1 - \alpha_0)^{-1} \psi, \quad t = 2, \ldots, T - 1, \tag{45}
\]

and

\[
E(y_{iT} \Delta y_{it}) = \alpha_0^{T-t-1}(1 - \alpha_0)^{-1} \psi + \alpha_0^{T-t} \Upsilon_t, \quad t = 2, \ldots, T - 1. \tag{46}
\]

Plugging into (14) and multiplying through by \((1 - \alpha_0)\psi^{-1}\) yields

\[
\alpha_0^{T-t-1} - (\alpha_0^{T-t-2} + \alpha_0^{T-t-2}) \alpha + \alpha_0^{T-t-2} \alpha^2 = 0, \quad t = 2, \ldots, T - 2, \tag{47a}
\]

\[
(\alpha_0^{T-2} + \alpha_0 \Sigma) - (\alpha_0^{T-2} + \alpha_0^{T-3} + \alpha_0^{T-3}) \alpha + \alpha_0^{T-3} \alpha^2 = 0, \tag{47b}
\]
where $\Sigma = \alpha_0^{T-t-1}(1-\alpha_0)\Psi^{-1}Y_t = (1-\alpha_0)\Psi^{-1}E(\nu_{iT-1}^2) - \alpha_0^{T-1}$ and where again we use the convention $0^0 = 1$.

If $T = 3$, then (47a) is empty and (47b) has roots $\alpha = \alpha_0$ and $\alpha = (1-\alpha_0^2) + E(\nu_{i2}^2)(1-\alpha_0)^{-1}$. Therefore, if the second root is not in $[0,1] \setminus \{\alpha_0\}$, then $\alpha_0$ is uniquely identified; otherwise $\alpha_0$ is partially identified. If $\alpha_0 = 0$, the second root can be written in terms of basic parameters as $\alpha = 1 + E(\nu_{i2}^2)/(E(c_i^2) - E(y_{i0}c_i))$. If $0 < \alpha_0 < 1$, the second root can be written as $\alpha = (1-\alpha_0^2) + \alpha_0 E(\nu_{i2}^2)/E(\nu_{i1}^2)$.

If $T \geq 4$, then (47a) and (47b) have two common roots if and only if $\Sigma = 0$. The latter is equivalent to

$$
\alpha_0^{T-1}\Psi + (\alpha_0 - 1)E(\nu_{iT-1}^2) = 0,
$$

which emphasizes the parallel with (13). The roots are $\alpha = \alpha_0$ and $\alpha = 1$.

### A.4 Proof of Theorem 4

By Assumption 1 and Equation (30), $E(\Delta y_{i1}) = 0$ is a necessary and sufficient condition for mean stationarity. From (31), this is equivalent to (23).

### B Stationarity and level moment restrictions

This appendix considers the role of the stationarity assumption (27) and the identification provided by the linear AB moment restrictions (5), the quadratic AS moment restrictions (6), and the level moment restrictions (28).

We begin by showing that under Assumptions 1 and 2 stationarity holds if and only if $\Psi = 0$. First, note that stationarity holds if and only if $E(c_i \Delta y_{i1}) = 0$. For the “if” part, note that $E(c_i \Delta y_{i-1}) = \alpha_0^{t-2}E(c_i \Delta y_{i1})$ for any $t = 2, \ldots, T$. It follows that $E(c_i \Delta y_{i1}) = 0$ implies $E(c_i \Delta y_{i-1}) = 0$ for $t = 2, \ldots, T$. For the “only if” part, take $t = 2$ in (27). Second, it is straightforward to verify that $\Psi = E(c_i \Delta y_{i1})$.

Theorem 5 shows that if the stationarity assumption holds, then the linear AB and the quadratic AS moment restrictions nearly identify $\alpha_0$. In particular, $\alpha_0$ is identified by the
linear AB and quadratic AS moment restrictions if $E(v_{it}^2) > 0$ for some $t = 1, \ldots, T-1$.

**Theorem 5** Suppose Assumptions 1 and 2 hold, the stationarity assumption (27) holds, and $T \geq 3$. Then linear AB and quadratic AS moment restrictions uniquely identify $\alpha_0$, except in the following case.

Any value of $\alpha$ solves the moment restrictions (full unidentification) if and only if $E(v_{it}^2) = 0$ for all $t = 1, \ldots, T-1$, (12) in Theorem 1 holds (i.e. $(\alpha_0 - 1)E(y_{i0}^2) + E(y_{i0}c_i) = 0$), and $\Psi = 0$ (i.e. $(\alpha_0 - 1)E(c_iy_{i0}) + E(c_i^2) = 0$).

**Proof** The conclusion follows from taking $\Psi = 0$ in Theorem 3. When $\alpha_0 = 1$, $E(c_i^2) = 0$, $E(v_{i0}^2) \neq 0$, and $E(v_{it}^2) = \lambda^{t-1}E(v_{i1}^2)$ for all $t = 2, \ldots, T-1$ and some $0 \leq \lambda < 1$, then there are two candidate solutions, namely $\alpha = \alpha_0$ and $\alpha = \lambda$ (partial identification). However, in this case it is known that $\alpha_0 = 1$.

If the stationarity assumption holds in addition to Assumptions 1 and 2, then the level moment restrictions are satisfied for $\alpha = \alpha_0$ (see e.g. Arellano and Bover, 1995). Theorem 6 shows that if the stationarity assumption holds, then the level moment restrictions do not provide identification over and above the linear AB and quadratic AS moment restrictions. (Of course they can contribute to efficiency.)

**Theorem 6** Suppose Assumptions 1 and 2 hold, the stationarity assumption (27) holds, and the linear AB moment restrictions do not identify $\alpha_0$. Then the level moment restrictions provide identification if and only if $E(v_{it}^2) > 0$ for some $t = 1, \ldots, T-1$.

**Proof** From (28), the level moment restrictions identify $\alpha_0$ whenever $E(y_{it-1}\Delta y_{it-1}) \neq 0$ for some $t = 2, \ldots, T$. When the linear AB moment restrictions do not identify $\alpha_0$ we have $E(y_{it-1}\Delta y_{it}) = 0$ for all $t = 1, \ldots, T-1$. The conclusion therefore follows from $E(y_{it}\Delta y_{it}) = \alpha_0E(y_{it-1}\Delta y_{it}) + E(v_{it}^2) = E(v_{it}^2)$ for $t = 1, \ldots, T-1$, which implies that $E(y_{it-1}\Delta y_{it-1}) \neq 0$ if and only if $E(v_{it}^2) > 0$.

Theorem 7 shows that the level moment restrictions can be misleading if the stationarity assumption does not hold.
Theorem 7 Suppose Assumptions 1 and 2 hold, the stationarity assumption (27) does not hold, and the linear AB moment restrictions do not identify \( \alpha_0 \). Then \( \alpha_0 \neq 1 \). Furthermore, if \( \mathbb{E}(v_{it-1}^2) = \alpha_0^{T-1}(1-\alpha_0)^{-1}\Psi \) holds, then \( \alpha = 1 \) is the unique solution; if not, then there are no common solution to the level moment restrictions.

Proof It is clear from (13) that \( \alpha_0 = 1 \) is not possible when \( \Psi \neq 1 \).

Define \( \tilde{u}_{it} = c_i + (\alpha_0 - 1)y_{it-1} + v_{it} \) and note that \( y_{it} = y_{it-1} + \tilde{u}_{it} \) for \( t = 2, \ldots, T \), so the level moment restrictions (28) are satisfied for \( \alpha = 1 \) if and only if \( \mathbb{E}(\tilde{u}_{it} \Delta y_{it-1}) = 0 \) for \( t = 2, \ldots, T - 1 \). By Assumption 2, we have

\[
\mathbb{E}(\tilde{u}_{it} \Delta y_{it-1}) = \mathbb{E}(c_i \Delta y_{it-1}) + (\alpha_0 - 1)\mathbb{E}(y_{it-1} \Delta y_{it-1})
\]

\[
= \mathbb{E}(c_i \Delta y_{it-1}) + (\alpha_0 - 1)\mathbb{E}(y_{it-1} \Delta y_{it-1}) + \mathbb{E}(c_i \Delta y_{it-1}) + \mathbb{E}(v_{it-1} \Delta y_{it-1}),
\]

\[t = 2, \ldots, T.\]

Since the linear AB moment restrictions do not identify \( \alpha_0 \), by (7) we have \( \mathbb{E}(y_{it-2} \Delta y_{it-1}) = 0 \) for \( t = 2, \ldots, T \). By Assumption 2, it follows that

\[
\mathbb{E}(\tilde{u}_{it} \Delta y_{it-1}) = \mathbb{E}(c_i \Delta y_{it-1}) + (\alpha_0 - 1)\mathbb{E}(y_{it-1} \Delta y_{it-1})
\]

\[
= \alpha_0 \mathbb{E}(c_i \Delta y_{it-1}) + (\alpha_0 - 1)\mathbb{E}(v_{it-1}^2), \quad t = 2, \ldots, T.
\]

Since \( \mathbb{E}(c_i \Delta y_{it-1}) = \alpha_0^{t-2}\mathbb{E}(c_i \Delta y_{i1}) = \alpha_0^{t-2}\Psi \) for \( t = 1, \ldots, T \) by Assumption 2 and the definition of \( \Psi \), we have

\[
\mathbb{E}(\tilde{u}_{it} \Delta y_{it-1}) = \alpha_0^{t-1}\Psi + (\alpha_0 - 1)\mathbb{E}(v_{it-1}^2), \quad t = 2, \ldots, T.
\]

By (13) in Theorem 1, we have \( \mathbb{E}(\tilde{u}_{it} \Delta y_{it-1}) = 0 \) for \( t = 2, \ldots, T - 1 \). It follows that if \( \mathbb{E}(\tilde{u}_{iT} \Delta y_{iT-1}) = 0 \), then \( \alpha = 1 \) is the unique solution to the level moment restrictions. If \( \mathbb{E}(\tilde{u}_{iT} \Delta y_{iT-1}) \neq 0 \), then there is no common solution to the level moment restrictions. \( \blacksquare \)
References


Hayakawa, K. (2009). On the effect of nonstationary initial conditions in dynamic panel...

<table>
<thead>
<tr>
<th>Group of DGPs</th>
<th>Conditions for LMR Identification Failure</th>
<th>Identification by QMR when LMR Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E(v_{it}^2) \neq 0 ) for some ( t = 1, \ldots, T-2 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Psi \neq 0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_0 = 0 )</td>
<td>( - )</td>
<td>( - )</td>
</tr>
<tr>
<td>( \alpha_0 = 1 )</td>
<td>( - )</td>
<td>( - )</td>
</tr>
<tr>
<td>( \alpha_0 \neq 0, \alpha_0 \neq 1 )</td>
<td>( (1 - \alpha_0)E(y_{i0}^2) = E(y_{i0}c_i) ) ( \text{and } E(v_{it}^2) = \alpha_0^t(1 - \alpha_0)^{-1}\Psi ) ( \text{for all } t = 1, \ldots, T-2 )</td>
<td>( T = 3: \text{Roots at } \alpha_0 ) and ( (1 - \alpha_0^2) + \alpha_0 E(v_{i2}^2)/E(v_{i1}^2) ); ( T \geq 4: \text{Unique if } E(v_{iT-1}^2) \neq \alpha_0^{T-1}(1 - \alpha_0)^{-1}\Psi, \text{else roots at } \alpha_0 ) and 1</td>
</tr>
<tr>
<td>( \Psi = 0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_0 = 0 )</td>
<td>( - )</td>
<td>( - )</td>
</tr>
<tr>
<td>( \alpha_0 = 1 )</td>
<td>( \text{E}(c_i^2) = 0 )</td>
<td>( T = 3: \text{Roots at } \alpha_0 ) and ( \text{E}(v_{i2}^2)/\text{E}(v_{i1}^2) ); ( T \geq 4: \text{Unique if no } \lambda \neq 1 \text{ such that } \text{E}(v_{it}^2) = \lambda\text{E}(v_{iT-1}^2) ) ( \text{for all } t = 2, \ldots, T-1, \text{else roots at } \alpha_0 ) and ( \lambda )</td>
</tr>
<tr>
<td>( \alpha_0 \neq 0, \alpha_0 \neq 1 )</td>
<td>( - )</td>
<td>( - )</td>
</tr>
<tr>
<td>( E(v_{it}^2) = 0 ) for all ( t = 1, \ldots, T-2 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Psi \neq 0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_0 = 0 )</td>
<td>( \text{E}(y_{i0}^2) = \text{E}(y_{i0}c_i) )</td>
<td>( T = 3: \text{Roots at } \alpha_0 ) and ( 1 + \text{E}(v_{i2}^2)/(\text{E}(c_i^2) - \text{E}(y_{i0}c_i)) ); ( T \geq 4: \text{Unique if } \text{E}(v_{iT-1}^2) \neq 0, \text{else roots at } \alpha_0 ) and 1</td>
</tr>
<tr>
<td>( \alpha_0 = 1 )</td>
<td>( - )</td>
<td>( - )</td>
</tr>
<tr>
<td>( \alpha_0 \neq 0, \alpha_0 \neq 1 )</td>
<td>( - )</td>
<td>( - )</td>
</tr>
<tr>
<td>( \Psi = 0 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_0 = 0 )</td>
<td>( \text{E}(y_{i0}^2) = \text{E}(y_{i0}c_i) )</td>
<td>( \text{Unique if } \text{E}(v_{iT-1}^2) \neq 0, \text{else no information} )</td>
</tr>
<tr>
<td>( \alpha_0 = 1 )</td>
<td>( \text{E}(c_i^2) = 0 )</td>
<td>( \text{Unique if } \text{E}(v_{iT-1}^2) \neq 0, \text{else no information} )</td>
</tr>
<tr>
<td>( \alpha_0 \neq 0, \alpha_0 \neq 1 )</td>
<td>( (1 - \alpha_0)\text{E}(y_{i0}^2) = \text{E}(y_{i0}c_i) )</td>
<td>( \text{Unique if } \text{E}(v_{iT-1}^2) \neq 0, \text{else no information} )</td>
</tr>
</tbody>
</table>

LMR: linear second moment restrictions; QMR: quadratic second moment restrictions; a dash indicates LMR provide identification for the entire group of DGPs.
Table 2: Comparison of GMM estimators for DGP 1: $\alpha_0 = 0.5$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Bias</th>
<th>SD</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DGP 1.1: first moment restrictions do not identify $\alpha_0$, $n = 200$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMR (5)</td>
<td>$-0.416$</td>
<td>$0.442$</td>
<td>$0.607$</td>
<td>$0.494$</td>
</tr>
<tr>
<td>LMR (5) + QMR (6)</td>
<td>$0.000$</td>
<td>$0.070$</td>
<td>$0.070$</td>
<td>$0.046$</td>
</tr>
<tr>
<td>LMR (5) with transformed data$^a$</td>
<td>$-0.404$</td>
<td>$0.440$</td>
<td>$0.597$</td>
<td>$0.485$</td>
</tr>
<tr>
<td>FMR (21) + LMR (5)</td>
<td>$-0.365$</td>
<td>$0.337$</td>
<td>$0.497$</td>
<td>$0.412$</td>
</tr>
<tr>
<td>FMR (21) + LMR (5) + QMR (6)</td>
<td>$-0.004$</td>
<td>$0.066$</td>
<td>$0.067$</td>
<td>$0.045$</td>
</tr>
<tr>
<td><strong>DGP 1.2: first moment restrictions do not identify $\alpha_0$, $n = 5000$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMR (5)</td>
<td>$-0.421$</td>
<td>$0.438$</td>
<td>$0.607$</td>
<td>$0.491$</td>
</tr>
<tr>
<td>LMR (5) + QMR (6)</td>
<td>$0.000$</td>
<td>$0.011$</td>
<td>$0.011$</td>
<td>$0.009$</td>
</tr>
<tr>
<td>LMR (5) with transformed data$^a$</td>
<td>$-0.417$</td>
<td>$0.436$</td>
<td>$0.603$</td>
<td>$0.490$</td>
</tr>
<tr>
<td>FMR (21) + LMR (5)</td>
<td>$-0.375$</td>
<td>$0.341$</td>
<td>$0.507$</td>
<td>$0.416$</td>
</tr>
<tr>
<td>FMR (21) + LMR (5) + QMR (6)</td>
<td>$0.000$</td>
<td>$0.011$</td>
<td>$0.011$</td>
<td>$0.009$</td>
</tr>
<tr>
<td><strong>DGP 1.3: first moment restrictions identify $\alpha_0$, $n = 200$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMR (5)</td>
<td>$-0.415$</td>
<td>$0.414$</td>
<td>$0.586$</td>
<td>$0.484$</td>
</tr>
<tr>
<td>LMR (5) + QMR (6)</td>
<td>$0.010$</td>
<td>$0.104$</td>
<td>$0.104$</td>
<td>$0.055$</td>
</tr>
<tr>
<td>LMR (5) with transformed data$^a$</td>
<td>$-0.049$</td>
<td>$0.147$</td>
<td>$0.155$</td>
<td>$0.122$</td>
</tr>
<tr>
<td>FMR (21) + LMR (5)</td>
<td>$-0.024$</td>
<td>$0.082$</td>
<td>$0.085$</td>
<td>$0.067$</td>
</tr>
<tr>
<td>FMR (21) + LMR (5) + QMR (6)</td>
<td>$-0.006$</td>
<td>$0.050$</td>
<td>$0.051$</td>
<td>$0.041$</td>
</tr>
</tbody>
</table>

SD: standard deviation; RMSE: root mean square error; MAE: mean absolute error; FMR: first moment restrictions; LMR: linear AB moment restrictions; QMR: quadratic AS moment restrictions.

$^a$All $y_{it}$ are transformed by adding 5. All simulations have 5000 Monte Carlo replications and use the optimal weight matrix. All experiments have $T = 4$ and $y_{i0}$, $c_i$, and $v_{i1}, \ldots, v_{iT}$ normally distributed. DGP 1.1 and 1.2: $\alpha_0 = 0.5$, $E(y_{i0}) = 0$, $E(c_i) = 0$, $E(y_{i0}^2) = 0.04$, $E(c_i^2) = 1$, $E(y_{i0}c_i) = 0.02$, $E(v_{i1}^2) = 0.99$, $E(v_{i2}^2) = 0.495$, $E(v_{i3}^2) = 1$, and $E(v_{i4}^2) = 1$. DGP 1.3: $\alpha_0 = 0.5$, $E(y_{i0}) = 1$, $E(c_i) = 1$, $E(y_{i0}^2) = 1$, $E(c_i^2) = 1$, $E(y_{i0}c_i) = 0$, $E(v_{i1}^2) = 3/2$, $E(v_{i2}^2) = 3/4$, $E(v_{i3}^2) = 1$, and $E(v_{i4}^2) = 1$. 

Table 3: Comparison of GMM estimators for DGP 2: $\alpha_0 = 1$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Bias</th>
<th>SD</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DGP 2.1: $E(v_{it}^2) = 1$ for all $t$, $n = 200$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMR (5)</td>
<td>-0.762</td>
<td>0.473</td>
<td>0.897</td>
<td>0.789</td>
</tr>
<tr>
<td>LMR (5) + QMR (6)</td>
<td>-0.108</td>
<td>0.114</td>
<td>0.157</td>
<td>0.133</td>
</tr>
<tr>
<td>LMR (5) with transformed data&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.766</td>
<td>0.469</td>
<td>0.898</td>
<td>0.791</td>
</tr>
<tr>
<td>FMR (21) + LMR (5)</td>
<td>-0.715</td>
<td>0.366</td>
<td>0.803</td>
<td>0.723</td>
</tr>
<tr>
<td>FMR (21) + LMR (5) + QMR (6)</td>
<td>-0.128</td>
<td>0.103</td>
<td>0.165</td>
<td>0.142</td>
</tr>
<tr>
<td><strong>DGP 2.2: $E(v_{it}^2) = 1$ for all $t$, $n = 5000$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMR (5)</td>
<td>-0.768</td>
<td>0.482</td>
<td>0.907</td>
<td>0.794</td>
</tr>
<tr>
<td>LMR (5) + QMR (6)</td>
<td>-0.041</td>
<td>0.049</td>
<td>0.064</td>
<td>0.052</td>
</tr>
<tr>
<td>LMR (5) with transformed data&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.773</td>
<td>0.471</td>
<td>0.905</td>
<td>0.793</td>
</tr>
<tr>
<td>FMR (21) + LMR (5)</td>
<td>-0.719</td>
<td>0.374</td>
<td>0.810</td>
<td>0.727</td>
</tr>
<tr>
<td>FMR (21) + LMR (5) + QMR (6)</td>
<td>-0.050</td>
<td>0.045</td>
<td>0.067</td>
<td>0.056</td>
</tr>
<tr>
<td><strong>DGP 2.3: $E(v_{12}^2) = 4$, $E(v_{it}^2) = 1$ for $t \neq 2$, $n = 200$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMR (5)</td>
<td>-0.661</td>
<td>0.444</td>
<td>0.796</td>
<td>0.683</td>
</tr>
<tr>
<td>LMR (5) + QMR (6)</td>
<td>-0.005</td>
<td>0.036</td>
<td>0.037</td>
<td>0.029</td>
</tr>
<tr>
<td>LMR (5) with transformed data&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.668</td>
<td>0.442</td>
<td>0.801</td>
<td>0.688</td>
</tr>
<tr>
<td>FMR (21) + LMR (5)</td>
<td>-0.553</td>
<td>0.339</td>
<td>0.648</td>
<td>0.562</td>
</tr>
<tr>
<td>FMR (21) + LMR (5) + QMR (6)</td>
<td>-0.007</td>
<td>0.036</td>
<td>0.037</td>
<td>0.029</td>
</tr>
</tbody>
</table>

SD: standard deviation; RMSE: root mean square error; MAE: mean absolute error; FMR: first moment restrictions; LMR: linear AB moment restrictions; QMR: quadratic AS moment restrictions.

<sup>a</sup>All $y_{it}$ are transformed by adding 5. All simulations have 5000 Monte Carlo replications and use the optimal weight matrix. All experiments have $T = 4$ and $y_{i0}$, $c_i$, and $v_{i1}, \ldots, v_{iT}$ normally distributed; $\alpha_0 = 1$, $E(y_{i0}) = 0$, $E(c_i) = 0$, $E(y_{i0}^2) = 1$, $E(c_i^2) = 0$, and $E(y_{i0}c_i) = 0$. 
When the linear AB moment restrictions do not identify $a_0$, entries in the same shading have the same value (except white), while entries in different shadings may or may not have the same value.

Figure 1: Matrix of second moments $[E(y_t y_{t+1})]$
RMSE: root mean square error; FMR: first moment restrictions; LMR: linear AB moment restrictions; QMR: quadratic AS moment restrictions. All parameters as in DGP 1.2, see Table 2, except $\alpha_0$ which varies along the horizontal axis.

Figure 2: RSME of GMM estimators near DGP 1.2
RMSE: root mean square error; FMR: first moment restrictions; LMR: linear AB moment restrictions; QMR: quadratic AS moment restrictions. All parameters as in DGP 2.2, see Table 3, except $\alpha_0$ which varies along the horizontal axis.

Figure 3: RSME of GMM estimators near DGP 2.2
RMSE: root mean square error; FMR: first moment restrictions; LMR: linear AB moment restrictions; QMR: quadratic AS moment restrictions. See Section 7.3 for details of the data generating processes.

Figure 4: RSME of GMM estimators near DGP 3