

Department of Economics

GMM Estimation and Inference in Dynamic Panel Data Models
with Persistent Data

Hugo Kruiniger

Working Paper No. 560

April 2006

ISSN 1473-0278



GMM estimation and inference in dynamic panel data models with persistent data

Hugo Kruiniger*
Queen Mary, University of London

This version: March 2006
(previous version: December 2003)

JEL classification: C12, C13, C23.

Keywords: dynamic panel data, GMM, weak instruments, weak identification, local asymptotics, multi-index asymptotics, diagonal path asymptotics, LM test, panel unit root test.

*Address: h.kruiniger@qmw.ac.uk; Dept. of Economics, Mile End Road, London E1 4NS.
I thank Steve Bond and Frank Windmeijer for kindly making one of their computer programs available to me. This research was funded by the ESRC under grant R000239139.

Abstract

In this paper we consider GMM based estimation and inference for the panel AR(1) model when the data are persistent and the time dimension of the panel is fixed. We find that the nature of the weak instruments problem of the Arellano-Bond estimator depends on the distributional properties of the initial observations. Subsequently, we derive local asymptotic approximations to the finite sample distributions of the Arellano-Bond estimator and the System estimator, respectively, under a variety of distributional assumptions about the initial observations and discuss the implications of the results we obtain for doing inference. We also propose two LM type panel unit root tests.

1 Introduction

In this paper we consider GMM based estimation and inference for the panel AR(1) model $y_{i,t} = \rho y_{i,t-1} + (1 - \rho)\mu_i + \varepsilon_{i,t}$, $i = 1, \dots, N$ and $t = -S + 1, \dots, T$, when the autoregression parameter ρ is close to or equal to one. Throughout the paper we assume that $S \geq -1$, the first observations occur at $t = 1$, the time dimension of the panel, T , is fixed (and small) and that the cross-section dimension of the panel, N , is large. Among other things we derive local asymptotic approximations to the finite sample distributions of some well-known linear GMM estimators for this model under a variety of assumptions about the initial observations and discuss the implications of the results we obtain for doing inference. We also propose two LM type panel unit root tests.

Persistent data raise at least three issues in connection with GMM estimation of the panel AR(1) model. First, some of the available moment conditions do not identify the unit root. For instance, it is well known that the GMM estimator of Arellano and Bond (1991), henceforth abbreviated as AB, breaks down when $\rho = 1$, because the instruments, which are lagged levels of the data ($y_{i,s}$, $s \leq t - 2$), are no longer correlated with the first-differences of the regressors ($\Delta y_{i,t-1}$): when $\rho = 1$, $E(y_{i,s} \Delta y_{i,t-1}) = E(y_{i,s} \varepsilon_{i,t-1}) = 0$. However, moment conditions such as $E(\Delta y_{i,s} (y_{i,t} - \rho y_{i,t-1})) = 0$ with $s \leq t - 1$ still identify the unit root, see Arellano and Bover (1995), henceforth abbreviated as Arbov. Below we will refer to an estimator that only exploits such moment conditions as an Arbov estimator. Second, most moment conditions are weak in some sense when ρ is close to unity. The finite-sample distributions of the corresponding estimators may therefore differ substantially from the first-order fixed-parameter asymptotic distributions. The second order bias approximation for the AB estimator derived by Hahn et al. (2001) also becomes inaccurate when ρ is close to unity. Third, the first-order derivatives with respect to ρ of some of the moment conditions are discontinuous at $\rho = 1$: it is easily verified for $t \geq 2$ that $\lim_{\rho \uparrow 1} E(y_{i,1} \Delta y_{i,t}) = -\frac{1}{2} \text{Var}(\varepsilon_{i,t}) \equiv -\frac{1}{2} \sigma^2$ when the data are covariance stationary, while $E(y_{i,1} \Delta y_{i,t}) = 0$ when $\rho = 1$.

In the paper it is shown that the weakness of the moment conditions implied by the panel AR(1) model can manifest itself in different ways depending on the distributional properties of the initial observations. In some cases the first-order derivatives of the moment conditions with respect to ρ , i.e. the Jacobians, tend to zero when ρ approaches one. In other cases, the moment conditions are weak because the standard deviations of the first-order derivatives of the underlying moment functions with respect to ρ explode when ρ approaches one. In the first type of situation the ‘signal’ of the moment functions becomes weak, whereas in the second type of situation the overwhelming ‘noise’ of the moment functions drowns their signal when ρ approaches one (cf Han and Phillips, 2006). In both situations doing a form of local asymptotics affords a better approximation to the finite sample distribution of the corresponding GMM estimator than doing conventional first-order fixed-parameter asymptotics.

When S is fixed (and small) and when ρ approaches one according to $\rho = 1 - \lambda N^{-1/2}$ as N grows large, where $\lambda > 0$, both the means and the standard deviations of the cross-sectional

averages of the first-order derivatives of the AB moment functions with respect to ρ vanish at a \sqrt{N} rate. Thus in this case we can obtain approximations to the finite sample distributions of AB estimators by applying the local-to-zero asymptotics of Staiger and Stock (1997). On the other hand, when the data are covariance stationary, the standard deviations of the first-order derivatives of the AB moment functions with respect to ρ explode when ρ approaches one, whereas their means tend to a nonzero constant due to the fact that $Var(y_{i,t} - \mu_i) = \sigma^2/(1 - \rho^2)$. In this case we can obtain ‘local-to-unity’ asymptotic approximations to the finite sample distributions of AB estimators by assuming that ρ approaches one according to $\rho = 1 - \lambda N^{-1}$ as N grows large, where again $\lambda > 0$. Moreover, we show that in this case the Arbov estimators and the two-step ‘optimal’ System estimator (cf Arellano and Bover, 1995) have non-normal distributions as well under local-to-unity asymptotics although they are still consistent when $T > 3$.

In the paper we also derive for all $d \geq 0$ the local-to-nonidentification asymptotic distributions of the AB estimator under the assumption that both S and N grow large with $S/N^d \rightarrow c > 0$. We show that if ρ approaches one according to $\rho = 1 - \lambda N^{-g}$ as N grows large, where $\lambda > 0$ and $0 < g \leq 1$, and if $0 \leq d \leq g$, then $E(y_{i,s}\Delta y_{i,t-1}) = O(N^{d-g})$ and $[Var(y_{i,s}\Delta y_{i,t-1})]^{1/2} = O(N^{\frac{1}{2}d})$, where $s \leq t - 2$. Thus the strength of both the signal and the noise of the AB moment functions increases with d , that is with S . The value of g for which the local-to-nonidentification asymptotic distributions of AB estimators are obtained also increases with d from a value of $1/2$ for $d = 0$ to a maximum value of 1 for $d \geq 1$. The value of g is chosen in such a way that the means and the standard deviations of the cross-sectional averages of the first-order derivatives of the AB moment functions with respect to ρ either do not vanish or vanish at the same rate when ρ approaches one according to $\rho = 1 - \lambda N^{-g}$ as N grows large, where $\lambda > 0$. When $d \geq 1$ and $g = 1$, both the Arbov and the two-step ‘optimal’ System estimator have the same non-normal local asymptotic distributions as under covariance stationarity when $g = 1$.

In the paper we also show that under local (-to-nonidentification) asymptotics (1) the ‘optimal’ AB estimators are biased downwards, (2) the estimators of the optimal weight matrices for the Arbov estimator and the System estimator and their asymptotic standard errors are no longer consistent when $g = 1$ and the data are covariance stationary or $d \geq 1$, (3) the vector of averages of the Arbov moment functions and the vector of their first-order

derivatives with respect to ρ are uncorrelated, (4) the two-step ‘optimal’ System estimator can have a skewed distribution, and (5) the truly optimal Arbov and System estimators have the same distribution, which is symmetric. The first result explains the downward biases of the AB estimator reported in a Monte Carlo study contained in Bond and Blundell (1998). The other results help to explain the findings of a Monte Carlo study contained in Bond and Windmeijer (2002), namely that when the data are covariance stationary and persistent Wald tests based on the two-step ‘optimal’ System estimator have incorrect size, whereas LM tests which are based on System estimators that use a restricted estimator of the weight matrix that is optimal under the null, have correct size. This paper therefore offers a theoretical justification for using LM tests in the context of panel AR models. Finally the paper proposes two LM-type panel unit root tests and studies their properties.

The paper is organised as follows. In section 2 we review GMM estimation of the panel AR(1) model. In section 3 we derive local asymptotic approximations to the finite sample distributions of the AB, Arbov and System estimators under various assumptions about the initial conditions, i.e. under various asymptotic plans for S and N . Section 4 proposes two LM-type panel unit root tests and investigates their properties both analytically and through Monte Carlo experiments. Section 5 concludes. Proofs are given in the Appendix.

A few words on notation. We use the symbol \xrightarrow{d} to signify convergence in distribution, the symbol plim to signify convergence in probability, and the symbol $\xrightarrow{q.m.}$ to signify convergence in quadratic mean. To state multi-index asymptotic results we make use of the following notation (see also Phillips and Moon, 1999, for definitions of the underlying concepts): $\lim_{S,N \rightarrow \infty, \text{seq}} N^{-1} \sum_{i=1}^N X_{i,t} = X_t$ is equivalent to $\lim_{N \rightarrow \infty} \lim_{S \rightarrow \infty} N^{-1} \sum_{i=1}^N X_{i,t} = X_t$ while $N^{-1/2} \sum_{i=1}^N X_{i,t} \xrightarrow{d} X_t$ as $(S, N \rightarrow \infty)_{\text{seq}}$ signifies that $N^{-1/2} \sum_{i=1}^N X_{i,t}$ converges in distribution to X_t sequentially by letting S pass to infinity first and letting N pass to infinity subsequently. Sometimes we only write $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N X_{i,t} = X_t$ instead of $\lim_{S,N \rightarrow \infty, \text{seq}} N^{-1} \sum_{i=1}^N X_{i,t} = X_t$ when it is clear that S passes to infinity first, e.g. when we have assumed covariance stationarity. Finally, $\text{plim}_{N \rightarrow \infty, S/N \rightarrow c} N^{-1} \sum_{i=1}^N X_{i,t} = X_t$ is an example of a so-called diagonal path probability limit. In this example both S and N pass to infinity with S/N converging to some constant c . We also make use of indicator functions. For instance, $1\{d = 0\} = 1$ if $d = 0$ and $1\{d = 0\} = 0$ if $d \neq 0$. Finally, the abbreviation PDS denotes Positive Definite Symmetric.

2 The panel AR(1) model

Consider the panel AR(1) model with random effects:

$$y_{i,t} = \rho y_{i,t-1} + v_{i,t}, \quad (1)$$

$$v_{i,t} = \eta_i + \varepsilon_{i,t}, \text{ where } \eta_i = (1 - \rho)\mu_i, \quad (2)$$

for $i = 1, \dots, N$ and $t = -S + 1, \dots, T$. The starting date of the $\{y_{i,t}\}$, $t = -S$, need not coincide with the date of the initial observations on y , $t = 1$, that is, $-S \leq 1$; the number of ‘individuals’, N , is large while the number of observations on y per individual, T , is fixed. Moreover $-1 < \rho \leq 1$. Note that the model can be rewritten as $y_{i,t} - \mu_i = \rho(y_{i,t-1} - \mu_i) + \varepsilon_{i,t}$.

The $(T + S + 1)$ -vectors $(y_{i,-S} \dots y_{i,T})'$, $i = 1, \dots, N$, are assumed to be i.i.d.¹ The composite error terms, the $v_{i,t}$, satisfy the following standard assumptions (cf Ahn and Schmidt, 1995):²

$$E(\mu_i) = 0 \text{ and } E(\mu_i^2) = \sigma_\mu^2, \text{ for } i = 1, \dots, N, \quad (3)$$

$$E(\varepsilon_{i,t}) = 0, \text{ } E(\varepsilon_{i,t}\eta_i) = 0, \text{ and} \quad (4)$$

$$E(\varepsilon_{i,t}y_{i,-S}) = 0, \text{ for } i = 1, \dots, N \text{ and } t = -S + 1, \dots, T, \quad (5)$$

and

$$E(\varepsilon_{i,s}\varepsilon_{i,t}) = 0, \text{ for } i = 1, \dots, N \text{ and } s \neq t, \text{ } s, t = -S + 1, \dots, T. \quad (6)$$

For convenience we also assume that the idiosyncratic errors, the $\varepsilon_{i,t}$, are homoskedastic over time:

$$E(\varepsilon_{i,t}^2) = \sigma^2, \text{ for } i = 1, \dots, N \text{ and } t = -S + 1, \dots, T. \quad (7)$$

The initial conditions are given by $y_{i,-S} = \mu_i + (1 - \rho)\xi_{i,-S}$, for $i = 1, \dots, N$. Note that when $\rho = 1$, the initial conditions are equal to the individual effects, the μ_i . Finally, we assume that $E(\xi_{i,-S}^8) < \infty$, $E(\mu_i^4) < \infty$, $E(\varepsilon_{i,t}^4) = \kappa\sigma^4$ and $E(\varepsilon_{i,t}^8) < \infty$, for $i = 1, \dots, N$ and $t = -S + 1, \dots, T$, where κ is the kurtosis parameter.

¹We assume identical distributions across the individuals for convenience.

²Note that $E(\varepsilon_{i,t}y_{i,s}) = 0$ for all t and all $s < t$ and $E(\varepsilon_{i,t}\varepsilon_{i,s}) = 0$ for all $t \neq s$ implies $E(\varepsilon_{i,t}\eta_i) = 0$ for all $t \geq -S + 2$ but not $E(\varepsilon_{i,-S+1}\eta_i) = 0$.

In the paper we consider various versions of the panel AR(1) model, which differ with respect to the assumptions made about the initial observations. Among them are the following two versions:

(CS) The $\{y_{i,t}\}$ have reached Covariance Stationarity at $t = 1$ when $|\rho| < 1$.

(FS) Fixed S : thus the $\{y_{i,t}\}$ have not reached stationarity at $t = 1$ when $|\rho| < 1$.

Ahn and Schmidt (1997) have shown that given assumptions (3)-(7), the $\{y_{i,t}\}$ have reached covariance stationarity at $t = 1$ if and only if the initial observations satisfy the following assumptions:

$$E(y_{i,1} - \mu_i) = 0, \quad E[(y_{i,1} - \mu_i)\mu_i] = 0, \quad \text{and} \quad (8)$$

$$Var(y_{i,1} - \mu_i) = \frac{\sigma^2}{1 - \rho^2}, \quad \text{for } |\rho| < 1 \text{ and } i = 1, \dots, N. \quad (9)$$

Note that the FS model does not rule out that mean-stationarity, i.e. assumption (8), holds.

2.1 GMM estimators for the panel AR(1) model

Arellano and Bond (1991, AB) proposed an GMM estimator which exploits the following $m = (T - 1)(T - 2)/2$ linear moment conditions:

$$E[y_{i,t-s}(\Delta y_{i,t} - \rho \Delta y_{i,t-1})] = 0 \text{ for } s = 2, \dots, t - 1 \text{ and } t = 3, \dots, T, \quad (10)$$

where $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$. These moment conditions are implied by assumptions (4)-(6). Note that they do not identify the unit root because $E(y_{i,t-s} \Delta y_{i,t-1}) = E(y_{i,t-s} \varepsilon_{i,t-1}) = 0$ when $\rho = 1$.

Arellano and Bover (1995) noted that if mean-stationarity, i.e. assumption (8), holds as well, one can add $T - 2$ linear moment conditions to those in (10):

$$E[(y_{i,t} - \rho y_{i,t-1}) \Delta y_{i,t-1}] = 0 \text{ for } t = 3, \dots, T. \quad (11)$$

The latter moment conditions do identify the unit root because $E(y_{i,t-1} \Delta y_{i,t-1}) = E(y_{i,t-1} \varepsilon_{i,t-1}) = \sigma^2$ when $\rho = 1$.

A GMM estimator that exploits the moment conditions in both (10) and (11) is known as a System (SYS) estimator.

The set of moment conditions in (10) and (11) is equivalent to a set that contains $T - 2$ Arellano-Bond and m Arellano-Bover type moment conditions:

$$E[y_{i,1}(\Delta y_{i,t} - \rho \Delta y_{i,t-1})] = 0 \text{ for } t = 3, \dots, T, \quad (12)$$

and

$$E[(y_{i,t} - \rho y_{i,t-1})\Delta y_{i,s}] = 0 \text{ for } s = 2, \dots, t - 1 \text{ and } t = 3, \dots, T. \quad (13)$$

A GMM estimator that only exploits the latter m moment conditions will be referred to as an Arellano-Bover (Arbov) estimator.

Let $y_i^t = [y_{i,1} \dots y_{i,t}]$ and let $Z_i = \text{diag}(y_i^1, \dots, y_i^{T-2})$ be a $(T - 2) \times m$ block-diagonal matrix. Then we can write the set of AB moment conditions in (10) as $E(Z_i' \Delta v_i) = 0$ where $\Delta v_i = v_i - v_{i,-1} = [\Delta v_{i,3} \dots \Delta v_{i,T}]'$. Under our assumptions, $E(\Delta v_i \Delta v_i') / \sigma^2 = H$, where $H = H_{T-2}$ is a $(T - 2)$ band-diagonal matrix with 2's on the main diagonal, -1 's on the first sub- and superdiagonal and zeros elsewhere. It follows that the AB GMM estimator which uses $W_{N,AB1} = (N^{-1} \sum_{i=1}^N Z_i' H Z_i)^{-1}$ as weight matrix is an optimal one-step GMM estimator. This estimator is denoted as $\hat{\rho}_{AB1}$. An AB estimator with an arbitrary weight matrix is simply denoted as $\hat{\rho}_{AB}$.

There exist no feasible optimal one-step weight matrices for the Arbov and SYS estimators, except when $\sigma_\eta^2 = 0$. Let $\Delta y_i^t = [\Delta y_{i,2} \dots \Delta y_{i,t}]$, let $Z_i^I = y_{i,1} I_{T-2}$, where I_{T-2} is an $(T - 2)$ identity matrix, let $Z_i^{II} = \text{diag}(\Delta y_i^2, \dots, \Delta y_i^{T-1})$ be a $(T - 2) \times m$ block-diagonal matrix and let $Z_i^S = \text{diag}(Z_i^I, Z_i^{II})$ be a $2(T - 2) \times (T - 2 + m)$ block-diagonal matrix. When $\sigma_\eta^2 = 0$, optimal one-step weight matrices for the Arbov and SYS estimators are given by $W_{N,Arbov1} = (N^{-1} \sum_{i=1}^N Z_i^{II'} Z_i^{II})^{-1}$ and $W_{N,SYS1} = (N^{-1} \sum_{i=1}^N Z_i^{S'} A Z_i^S)^{-1}$, respectively, where A is given by

$$A = \begin{bmatrix} H & C \\ C' & I_{T-2} \end{bmatrix},$$

where $C = C_{T-2}$ is a $(T - 2) \times (T - 2)$ matrix with ones on the main diagonal, -1 's on the first subdiagonal and zeros elsewhere. Note that $C = E(v_i \Delta v_i') / \sigma^2$. The one-step GMM estimators based on $W_{N,Arbov1}$ and $W_{N,SYS1}$ will be referred to as the Arbov1 estimator and the SYS1 estimator, respectively.

The optimal two-step System (SYS2) estimator is based on the weight matrix $W_{N,SYS2}(\hat{\rho}_1) = (N^{-1} \sum_{i=1}^N Z_i^{S'} [\widehat{\Delta v}_i' \widehat{v}_i'] [\widehat{\Delta v}_i \widehat{v}_i] Z_i^S)^{-1}$, where $\widehat{\Delta v}_i = \Delta y_i - \hat{\rho}_1 \Delta y_{i,-1}$ and

$\widehat{v}_i = y_i - \widehat{\rho}_1 y_{i-1}$ with $\widehat{\rho}_1$ an initial consistent estimator for ρ . We define the optimal two-step AB (AB2) and Arbov (Arbov2) estimators and their weight matrices analogously. The $W_{N,SYSk}$ matrices ($k = 1, 2$) can be partitioned as

$$W_{N,SYSk} = \begin{bmatrix} W_{N,SYSk,11} & W_{N,SYSk,12} \\ W_{N,SYSk,21} & W_{N,SYSk,22} \end{bmatrix},$$

where the $W_{N,SYSk,11}$ block corresponds to $E(Z_i' \Delta v_i) = 0$.

Let $\widetilde{d}_t = \text{diag}(0, \dots, 0, 1, \dots, 1)$ be a diagonal matrix with t zeros and $T - t - 2$ ones on the diagonal and let d_t be a matrix that comprises the last $T - t - 2$ columns of \widetilde{d}_t . In addition, let $Z_i^{AB} = [y_{i,1} I_{T-2} \ Z_i^D] = [Z_i^I \ Z_i^D]$, where $Z_i^D = [d_1 \Delta y_{i,2} \ d_2 \Delta y_{i,3} \ \dots \ d_{T-3} \Delta y_{i,T-2}]$ is a $(T-2) \times [m - (T-2)]$ matrix. Then we can rewrite the set of m AB moment conditions in $E(Z_i' \Delta v_i) = 0$ as $E(Z_i^{AB'} \Delta v_i) = 0$. Thus there exists a nonsingular constant matrix K^{AB} such that $Z_i' = K^{AB} Z_i^{AB'}$.

An alternative transformation that can be used for removing individual effects is the Helmert transformation which amounts to taking forward orthogonal deviations. The Helmert transformation of $(v_{i,2} \ v_i')'$ is $\widetilde{v}_i = (\widetilde{v}_{i,2} \dots \widetilde{v}_{i,T-1})'$, where $\widetilde{v}_{i,t} = \left(\frac{T-t}{T-t+1}\right)^{1/2} \left[v_{i,t} - \frac{1}{T-t} \sum_{s=t+1}^T v_{i,s} \right]$, $t = 2, \dots, T-1$. Note that $v_{i,t} - \frac{1}{T-t} \sum_{s=t+1}^T v_{i,s}$ is equal to a linear combination of first differences of the errors. An advantage of using the Helmert transformation rather than taking first differences is that it preserves the orthogonality among the errors, i.e. if $E(\varepsilon_i \varepsilon_i') = \sigma^2 I_{T-2}$, then $E(\widetilde{\varepsilon}_i \widetilde{\varepsilon}_i') = \sigma^2 I_{T-3}$.

Applying the Helmert transformation to $y_{i,t} = \rho y_{i,t-1} + (1 - \rho)\mu_i + \varepsilon_{i,t}$, $t = 2, \dots, T$, yields $\widetilde{y}_{i,t} = \rho \widetilde{y}_{i-1,t-1} + \widetilde{\varepsilon}_{i,t}$, $t = 2, \dots, T-1$, where $\widetilde{y}_{i-1,t-1} = \left(\frac{T-t}{T-t+1}\right)^{1/2} \left[y_{i,t-1} - \frac{1}{T-t} \sum_{s=t}^{T-1} y_{i,s} \right]$. Note that all lagged values of $y_{i,t}$ are valid instruments for the t -th equation of the transformed system. If we let $z'_{i,t} = y_i^{t-1}$, then the set of AB moment conditions in (10) is equivalent to the following set of moment conditions: $E(z_{i,t} \widetilde{v}_{i,t}) = 0$, $t = 2, \dots, T-1$. That is, there is a nonsingular matrix K^H such that $Z_i' \Delta v_i = K^H Z_i' \widetilde{v}_i$, see Arellano and Bover (1995). An optimal weight matrix for $E(Z_i' \widetilde{v}_i) = 0$ is $[N^{-1} \sum_{i=1}^N (Z_i' Z_i)]^{-1}$, which is a block-diagonal matrix with typical diagonal block equal to $[N^{-1} \sum_{i=1}^N (z_{i,t} z'_{i,t})]^{-1}$. Note that $Z_i' H Z_i = K^H Z_i' Z_i K^H'$. It follows that the one-step AB GMM estimator can be rewritten as a weighted average of $T-2$ 2SLS estimators: $\widehat{\rho}_{AB1} = \left\{ \sum_{t=2}^{T-1} [\widetilde{y}'_{-1,t-1} Z_t (Z_t' Z_t)^{-1} Z_t' \widetilde{y}_{-1,t-1}] \right\}^{-1} \sum_{t=2}^{T-1} [\widetilde{y}'_{-1,t-1} Z_t (Z_t' Z_t)^{-1} Z_t' \widetilde{y}_t]$ where $\widetilde{y}_t = [\widetilde{y}_{1,t} \ \dots \ \widetilde{y}_{N,t}]'$, $\widetilde{y}_{-1,t-1} = [\widetilde{y}_{1,-1,t-1} \ \dots \ \widetilde{y}_{N,-1,t-1}]'$ and $Z_t = [z_{1,t} \ \dots \ z_{N,t}]'$.

3 Asymptotic properties of the estimators when the data are persistent

Below we investigate the asymptotic properties of various AB, Arbov and System estimators when ρ is close to unity under three different asymptotic schemes: S is finite (the FS model), $S, N \rightarrow \infty$ sequentially with $S \rightarrow \infty$ first (the CS model), and finally $S, N \rightarrow \infty$ simultaneously with $S/N^d \rightarrow \hat{c} > 0$, where $d \geq 0$.

In the analysis the ratios $Var(y_{i,t} - \mu_i)/\sigma^2$ play an important role. The panel AR(1) model implies the following expression for $Var(y_{i,1} - \mu_i)$ when $|\rho| < 1$:

$$Var(y_{i,1} - \mu_i) = \rho^{2(1+S)}Var(y_{i,-S} - \mu_i) + (1 - \rho^{2(1+S)})\sigma^2/(1 - \rho^2). \quad (14)$$

It is easily verified for any S and $\rho > 0$ that $d[Var(y_{i,1} - \mu_i)]/d\rho > 0$ and $\lim_{\rho \uparrow 1} Var(y_{i,1} - \mu_i) = Var(y_{i,-S} - \mu_i) + (1 + S)\sigma^2$. Moreover, $\lim_{S \rightarrow \infty} Var(y_{i,1} - \mu_i) = \sigma^2/(1 - \rho^2)$. It follows that when S is large and ρ tends to one, the ratio $Var(y_{i,1} - \mu_i)/\sigma^2$ becomes large. Indeed, $\lim_{\rho \uparrow 1} \lim_{S \rightarrow \infty} Var(y_{i,1} - \mu_i) = \infty$. However, when S is finite $\lim_{\rho \uparrow 1} Var(y_{i,1} - \mu_i) < \infty$.

3.1 Local-to-zero asymptotics for the fixed S case

In this subsection we assume that S is finite. Then $\lim_{\rho \uparrow 1} Var(y_{i,t} - \mu_i) < \infty$ for $t = 1, \dots, T$.

Let $E[m_{AB,s,t}(\rho)] = 0$ with $m_{AB,s,t}(\rho) = y_{i,s}(\Delta y_{i,t} - \rho \Delta y_{i,t-1})$ and $s \leq t - 2$ represent an arbitrary AB moment condition from (10). Then it is easily verified that $E(dm_{AB,s,t}/d\rho) = \rho^{t-2-s}(1 - \rho)E[(y_{i,s} - \mu_i + \mu_i)(y_{i,s} - \mu_i)]$ and hence $\lim_{\rho \uparrow 1} E(dm_{AB,s,t}/d\rho) = 0$.

Consider now the simple first-stage regression $\Delta y_{i,t-1} = \pi y_{i,s} + \omega_i$, with $s \leq t - 2$, which corresponds to an arbitrary AB moment condition. Let $\hat{\pi} = (\sum_i y_{i,s}^2)^{-1} \sum_i y_{i,s} \Delta y_{i,t-1}$. Clearly, $\lim_{\rho \uparrow 1} \lim_{N \rightarrow \infty} \hat{\pi} = 0$. Moreover, if $\rho = 1 - \lambda N^{-1/2}$, $\pi = [E(y_{i,s}^2)]^{-1} E(y_{i,s} \Delta y_{i,t-1}) = O(N^{-1/2})$ and the first-stage F-statistic $F = \hat{\pi}^2 N (\sum_i y_{i,s}^2) / \sum_i (\Delta y_{i,t-1} - \hat{\pi} y_{i,s})^2 = O_p(1)$. Therefore, when ρ is close to unity, AB GMM estimators suffer from a standard weak instruments problem (cf Staiger and Stock, 1997). Considering the multiple first-stage regressions $\Delta y_{i,t-1} = \sum_{k=1}^{t-2} \pi_k y_{i,k} + \omega_i$, $t = 3, \dots, T$ leads to the same conclusions because T is fixed. ^{3 4}

³Instead one could consider the first-stage regressions corresponding to $E(z_{i,t} \tilde{\varepsilon}_{i,t}) = 0$, $t = 2, \dots, T - 1$, i.e. $\tilde{y}_{i,t-1} = \sum_{k=1}^{t-1} \tilde{\pi}_k y_{i,k} + \tilde{\omega}_i$, $t = 2, \dots, T - 1$, but that would lead to the same

Staiger and Stock have argued that doing local-to-zero asymptotics may provide a better approximation to the finite sample distribution of a GMM estimator that exploits weak moment conditions than traditional first-order fixed parameter asymptotics. In the context of the AB estimator for the FS version of the panel AR(1) model, doing local-to-zero asymptotics requires choosing a parameter sequence for ρ such that the first stage regression parameter π tends to zero as $N^{-1/2}$ when the sample size increases, i.e. $\pi = O(N^{-1/2})$. As we have seen above, in the FS version of the panel AR(1) model (where S is fixed), $\rho = 1 - \lambda N^{-1/2}$ implies that $\pi = O(N^{-1/2})$. The local-to-zero approach recognizes that for $\rho = 1 - \lambda N^{-1/2}$, $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N Z_i' \Delta y_{i,-1} = 0$ and that for this parameter sequence the mean of the vector $N^{-1/2} \sum_{i=1}^N Z_i' \Delta y_{i,-1}$ remains of the same order of magnitude as the standard deviations of its elements when N grows large.

We note that the Arellano-Bover estimator does not suffer from a weak instruments problem when S is fixed.⁵ We have the following result:

Theorem 1 *Let S be finite, let $\rho = 1 - \lambda N^{-1/2}$ with $\lambda \geq 0$, let $\sigma_{\mu y} = E(y_{i,1} \mu_i)$ and let $\sigma_y^2 = E(y_{i,1}^2)$. Let $\hat{\rho}_1$ be an initial \sqrt{N} -consistent estimator for ρ , i.e. $\sqrt{N}(\hat{\rho}_1 - \rho) \xrightarrow{d} \mathcal{K} \neq 0$. Let W_N be an arbitrary sequence of PDS weight matrices with $\text{plim}_{N \rightarrow \infty} W_N = W$, where W is PDS. Finally, let $W_{AB1} = \lim_{N \rightarrow \infty} [E(Z_i' H Z_i)]^{-1}$ and $W_{AB1b} = \lim_{N \rightarrow \infty} [E(Z_i' Z_i)]^{-1}$.*

$$(a1) \text{ If } T = 3, \text{ then } \hat{\rho}_{AB} - \rho \xrightarrow{d} \frac{\check{X}_1}{\check{X}_2} \text{ with } \begin{bmatrix} \check{X}_1 \\ \check{X}_2 \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ \lambda(\sigma_{\mu y} - \sigma_y^2) \end{pmatrix}, \sigma^2 \sigma_y^2 \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right].$$

$$(a2) \text{ If } T \geq 3, \text{ then } N^{-1/2} \sum_{i=1}^N Z_i' \Delta v_i \xrightarrow{d} \check{X}_5, N^{-1/2} \sum_{i=1}^N Z_i' \Delta y_{i,-1} \xrightarrow{d} \check{X}_6 \text{ and } \hat{\rho}_{AB} - \rho \xrightarrow{d}$$

conclusions.

⁴The bias of a 2SLS estimator that exploits $E(y_i^{t-2} \Delta \varepsilon_{i,t}) = 0$ is approximately equal to the ratio of $E(\omega_i \varepsilon_i) / E(\omega_i^2)$ and the expected value of a first-stage ‘F-statistic’, \bar{F} ; $E(\bar{F}) \approx 1 + \tau^2 / K$, where $\tau^2 = \pi' E[y_i^{t-2} (y_i^{t-2})'] \pi / E(\omega_i^2)$ is the so-called concentration parameter and $K = t - 2$ is the number of instruments. See e.g. Hahn and Hausman (2002). When $\rho = 1 - \lambda N^{-0.5}$, $\lim_{N \rightarrow \infty} E(\omega_i \Delta \varepsilon_{i,t}) / E(\omega_i^2) = E(\Delta y_{i,t-1} \Delta \varepsilon_{i,t}) / E[(\Delta y_{i,t-1})^2] = -1/2$ and $E(F) = O(1)$.

⁵Consider the simple ‘Arbov’ first-stage regression $y_{i,t-1} = \pi \Delta y_{i,s} + \omega_i$. Let $\hat{\pi} = [\sum_i (\Delta y_{i,s})^2]^{-1} \sum_i y_{i,t-1} \Delta y_{i,s}$. Clearly, $\lim_{r \uparrow 1} \text{plim}_{N \rightarrow \infty} \hat{\pi} = 1 \neq 0$. Moreover, the first stage F-statistic $F = \hat{\pi}^2 N \sum_i (\Delta y_{i,s})^2 / \sum_i (y_{i,t-1} - \hat{\pi} \Delta y_{i,s})^2 = O_p(N)$.

$\check{X}'_6 W \check{X}_5 / \check{X}'_6 W \check{X}_6$ with $\check{X}_5 \sim N(0, \check{\Sigma}_{55})$, $\check{X}_6 \sim N(\check{\mu}_6, \check{\Sigma}_{66})$ and $E(\check{X}_5 \check{X}'_6) = \check{\Sigma}_{56} \neq 0$, where $\check{\mu}_6 = \check{\mu}_6(\lambda)$, $\check{\Sigma}_{55}$, $\check{\Sigma}_{56}$ and $\check{\Sigma}_{66}$ are given in the proof; when $T = 3$, $\check{X}_5 = \check{X}_1$ and $\check{X}_6 = \check{X}_2$.

(b) If $T \geq 3$, then $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N Z_i^{II'} y_{i,-1} = \check{X}_8 = \sigma^2 \iota_m$, $N^{-1/2} \sum_{i=1}^N Z_i^{II'} v_i \xrightarrow{d} \check{X}_7$, $\text{plim}_{N \rightarrow \infty} (\hat{\rho}_{Arbov} - \rho) = 0$ and $\sqrt{N}(\hat{\rho}_{Arbov} - \rho) \xrightarrow{d} \check{X}'_8 W \check{X}_7 / \check{X}'_8 W \check{X}_8$, where ι_m is an m -vector of ones and $\check{X}_7 \sim N(0, \sigma^4 I_m)$.

(c) $\text{plim}_{N \rightarrow \infty} W_{N,AB1} = W_{AB1}$ is PDS and $\text{plim}_{N \rightarrow \infty} [N^{-1} \sum_{i=1}^N (Z_i' Z_i)]^{-1} = W_{AB1b}$ is PDS.

(d) $\text{plim}_{N \rightarrow \infty} W_{N,Arbov1} = \sigma^2 \times \text{plim}_{N \rightarrow \infty} W_{N,Arbov2}(\hat{\rho}_1) = \sigma^2 W_{Arbov2} \equiv \sigma^{-2} I_m$.

(e) $\text{plim}_{N \rightarrow \infty} W_{N,SYs1} = \sigma^2 \times \text{plim}_{N \rightarrow \infty} W_{N,SYs2}(\hat{\rho}_1) = \sigma^2 W_{SYs2} \equiv \text{diag}(\sigma_y^{-2} H^{-1}, \sigma^{-2} I_m)$.

(f) If $T \geq 3$, then $N^{-1/2} \sum_{i=1}^N Z_i^{II'} \Delta v_i \xrightarrow{d} \check{X}_9$, $\text{plim}_{N \rightarrow \infty} (\hat{\rho}_{SYs} - \rho) = 0$ and $\sqrt{N}(\hat{\rho}_{SYs} - \rho) \xrightarrow{d} (\check{X}'_8 W_{21} \check{X}_9 + \check{X}'_8 W_{22} \check{X}_7) / (\check{X}'_8 W_{22} \check{X}_8)$, where \check{X}_9 is a sub-vector of \check{X}_5 .

(g) $\text{Asyvar}(\hat{\rho}_{SYs2}) = \text{Asyvar}(\hat{\rho}_{Arbov2}) = 1/m$.

(h) If $T > 3$, then $\hat{\rho}_{AB1}$ is asymptotically biased downwards.

PROOF

See appendix A.1.

Theorem 1 implies that if $\rho = 1 - \lambda N^{-1/2}$, $\hat{\rho}_{AB1}$ is inconsistent, i.e. $\hat{\rho}_{AB1} - \rho \xrightarrow{d} \omega_{AB1}$ where $\omega_{AB1} = \check{X}'_6 W_{AB1} \check{X}_5 / \check{X}'_6 W_{AB1} \check{X}_6$. Similarly, $\hat{\rho}_{AB2}$ is inconsistent. Moreover, if $\rho = 1 - \lambda N^{-1/2}$, then $W_{N,AB2}^{-1}(\hat{\rho}_{AB1}) \xrightarrow{d} \omega_{AB1}^2 \lim_{\rho \uparrow 1} E(Z_i' \Delta y_{i,-1} \Delta y'_{i,-1} Z_i) - 2\omega_{AB1} \lim_{\rho \uparrow 1} E(Z_i' \Delta y_{i,-1} \Delta v'_i Z_i) + \sigma^2 W_{AB1}^{-1} \neq \sigma^2 W_{AB1}^{-1}$.

When $T = 3$, $\omega_{AB1} = \check{X}_1 / \check{X}_2$. The distribution of the ratio of two, possibly correlated, normal variables has been studied by Fieller (1932). This ratio does not have finite moments.

For $\lambda = 0$, one obtains the asymptotic distribution of an AB GMM estimator for $\rho = 1$. The local-to-zero asymptotic distribution of the AB GMM estimators also captures the fact that this estimator is biased downwards when $T > 3$ and ρ is close to unity. The bias results from the fact that the instruments are weak and the fact that $E(\check{X}_5 \check{X}'_6) \neq 0$.

When $\rho = 1 - \lambda N^{-1/2}$, $\hat{\rho}_{SYs1}$ and $\hat{\rho}_{SYs2}$ are consistent despite the fact that W_{SYs1} and W_{SYs2} are PDS and therefore give some weight to the weak AB moment conditions. The reason for these consistency results is that although $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N Z_i^{II'} \Delta y_{i,-1} = 0$, $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N Z_i^{II'} y_{i,-1} = \check{X}_8 \neq 0$. However, when $\rho = 1 - \lambda N^{-1/2}$ exploiting the AB moment conditions does not reduce the asymptotic variance of $\hat{\rho}_{SYs2}$, $\text{Asyvar}(\hat{\rho}_{SYs2})$.

3.2 Local-to-unity asymptotics for the covariance stationary case

We now assume that the CS version of the model applies, i.e. $S \rightarrow \infty$ and the model has reached stationarity at $t = 1$ when $|\rho| < 1$. In the CS model $\lim_{\rho \uparrow 1} \text{Var}(y_{i,t})/\sigma^2 = \lim_{\rho \uparrow 1} [\text{Var}(y_{i,t} - \mu_i)/\sigma^2 + \sigma_\mu^2/\sigma^2] = \lim_{\rho \uparrow 1} [1/(1 - \rho^2)] + \sigma_\mu^2/\sigma^2 = \infty$ for $t = 1, \dots, T$. Moreover, $E(-y_{i,s}\Delta y_{i,t-1}) = (1 - \rho)\rho^{t-2-s}E[(y_{i,s} - \mu_i + \mu_i)(y_{i,s} - \mu_i)] = \sigma^2(1 - \rho)\rho^{t-2-s}/(1 - \rho^2)$ when $s \leq t - 2$. It follows that $\lim_{\rho \uparrow 1} E(dm_{AB,s,t}/d\rho) = \sigma^2/2$. Since $E[(\Delta y_{i,t-1})^2] = 2\sigma^2/(1 + \rho)$, we also have $\lim_{\rho \uparrow 1} E(dm_{Arbov,s,t}/d\rho) = \lim_{\rho \uparrow 1} E(-y_{i,s}\Delta y_{i,t-1}) = -\sigma^2/2$ when $s > t - 2$. On the other hand, $E(y_{i,s}\Delta y_{i,t-1})|_{\rho=1} = E(y_{i,s}\varepsilon_{i,t-1}) = 0$ when $s \leq t - 2$, while $E(y_{i,s}\varepsilon_{i,t-1}) = \sigma^2$ when $s > t - 2$. This implies that the $E(y_{i,s}\Delta y_{i,t-1})$ are discontinuous at $\rho = 1$.

Consider again the simple first-stage regression $\Delta y_{i,t-1} = \pi y_{i,s} + \omega_i$ ($s \leq t - 2$), which corresponds to an arbitrary AB moment condition. Then we find again that $\lim_{\rho \uparrow 1} \text{plim}_{N \rightarrow \infty} \hat{\pi} = \lim_{\rho \uparrow 1} \text{plim}_{N \rightarrow \infty} (\sum_i y_{i,s}^2)^{-1} \sum_i y_{i,s} \Delta y_{i,t-1} = 0$. Moreover, if $\rho = 1 - \lambda N^{-1}$, the first-stage F-statistic $F = \hat{\pi}^2 N (\sum_i y_{i,s}^2) / \sum_i (\Delta y_{i,t-1} - \hat{\pi} y_{i,s})^2 = O_p(1)$ and $\pi = [E(y_{i,s}^2)]^{-1} E(y_{i,s} \Delta y_{i,t-1}) = O(N^{-1})$. Note that if one would choose the parameter sequence $\rho = 1 - \lambda N^{-1/2}$, then $\pi = O(N^{-1/2})$ but $F = O_p(N^{1/2})$. Therefore when the CS version of the model applies and ρ is close to unity, the AB GMM estimator still suffers from some sort of weak instruments problem (see also Blundell and Bond, 1998), albeit not from one of the Staiger-Stock type. In this case the problem arises because $\lim_{\rho \uparrow 1} E(y_{i,s}^2) \rightarrow \infty$, whereas $\lim_{\rho \uparrow 1} E(y_{i,s} \Delta y_{i,t-1}) = -\sigma^2/2 \neq 0$. Considering the multiple first-stage regressions $\Delta y_{i,t-1} = \sum_{k=1}^{t-2} \pi_k y_{i,k} + \omega_i$ for $t = 3, \dots, T$ still leads to the same conclusion because T is fixed.

Consider now the simple first-stage regression $y_{i,t-1} = \pi \Delta y_{i,s} + \omega_i$ ($s \leq t - 1$), which corresponds to an arbitrary Arellano-Bover moment condition. Clearly, $\lim_{\rho \uparrow 1} \text{plim}_{N \rightarrow \infty} \hat{\pi} = 1/2 \neq 0$. Nonetheless, if $\rho = 1 - \lambda N^{-1}$, the first-stage F-statistic $F = \hat{\pi}^2 N \sum_i (\Delta y_{i,s})^2 / \sum_i (y_{i,t-1} - \hat{\pi} \Delta y_{i,s})^2 = O_p(1)$, even though $\pi = O(1)$. The reason for this finding is that when $S \rightarrow \infty$, $E(y_{i,t-1}^2) \propto 1/(1 - \rho)$ and hence $\text{plim}_{N \rightarrow \infty} N^{-2} \sum_i (y_{i,t-1} - \hat{\pi} \Delta y_{i,s})^2 = \sigma^2/(2\lambda)$ if $\rho = 1 - \lambda N^{-1}$. However, $\lim_{\rho \uparrow 1} E(y_{i,t-1} \Delta y_{i,s}) = \sigma^2/2 \neq 0$.

Notwithstanding that the AB and the Arbov moment conditions are not weak in the traditional sense of Staiger and Stock (in the sense that $F = O_p(1)$ when $\pi = O(N^{-1/2})$), the finite sample distributions of the corresponding GMM estimators differ considerably from their first-order fixed-parameter asymptotic distributions when ρ is close to one. See, for

instance, the Monte Carlo evidence reported 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 ρ approaches one. For instance, in appendix A.1 we show that $Var(y_{i,1}\Delta y_{i,2}) \propto 1/(1-\rho)$, $Var(y_{i,2}\Delta y_{i,2}) \propto 1/(1-\rho)$, $Var(y_{i,1}\Delta\varepsilon_{i,3}) \propto 1/(1-\rho)$ and also $Cov(y_{i,1}\Delta\varepsilon_{i,3}, y_{i,1}\Delta y_{i,2}) \propto 1/(1-\rho)$, see lemma 9. We obtain local-to-unity approximations to the distributions of the estimators by choosing a parameter sequence such that the variances of the cross-products of the instruments and the regressors become $O(1)$, that is by choosing $\rho = 1 - \lambda N^{-1}$. For $T = 3$, we have the following local-to-unity asymptotic results for the AB and Arbov estimators:

Theorem 2 *Let the CS model hold, let $T = 3$ and let $\rho = 1 - \lambda N^{-1}$ with $\lambda > 0$. Then*

$$(a) \hat{\rho}_{AB} - \rho \xrightarrow{d} \frac{\tilde{X}_1}{\tilde{X}_2}, \quad \text{where } \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix} \sim N \left[-\frac{1}{2}\sqrt{\lambda} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right],$$

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

PROOF

See appendix A.2.

The local-to-unity asymptotic results of theorem 2 have been obtained under the assumption of covariance stationarity. However, if we take $\lambda = 0$ in the local-to-unity asymptotic distribution of the AB estimator, we obtain its distribution for $\rho = 1$, despite the fact that we need to condition on initial conditions in this case.

Note that the local parameter sequence that is used to derive the local-to-unity asymptotic distribution of the AB estimator is different from the one used in theorem 1. This is related to the fact that under covariance stationarity the second moments of the initial observations, which also appear in the local-to-zero distribution, are proportional to $1/(1-\rho)$.

We remark that the local-to-unity parameter sequences depend on N . Recently, Moon and Phillips (2000) have also considered estimation of autoregressive roots near unity using panel data. However, they considered consistent estimation procedures for the localizing parameter $c < 0$ in $\rho = \exp(c/T)$ assuming that T grows large.

Note that the set of m Arbov moment conditions in $E(Z_i^{II'}v_i) = 0$ can be restated as $E(\tilde{Z}_i^{II'}[\Delta v_i' v_i']') = 0$, where $\tilde{Z}_i^{II'} = \text{diag}(Z_i^D, Z_i^L)$ is a $2(T-2) \times m$ matrix with $Z_i^L = \text{diag}(\Delta y_{i,2}, \dots, \Delta y_{i,T-1})$. We have the following local-to-unity asymptotic results for $T > 3$:

Theorem 3 *Let the CS model hold, let $T > 3$ and let $\rho = 1 - \lambda N^{-1}$ with $\lambda > 0$. Let $\hat{\rho}_1$ be an initial \sqrt{N} -consistent estimator for ρ , i.e. $\sqrt{N}(\hat{\rho}_1 - \rho) \xrightarrow{d} \mathcal{K} \neq 0$. In addition, let K^{II} be the nonsingular constant matrix such that $\tilde{Z}_i^{II'}[\Delta v_i' v_i']' = K^{II} Z_i^{II'} v_i$. Finally, let W_N be an arbitrary sequence of PDS weight matrices with $\text{plim}_{N \rightarrow \infty} W_N = W$, where W is PDS. Then*

(a) $N^{-1} \sum_{i=1}^N Z_i^{I'} \Delta v_i \xrightarrow{d} \tilde{X}_{51}$, $N^{-1} \sum_{i=1}^N Z_i^{I'} \Delta y_{i,-1} \xrightarrow{d} \tilde{X}_{61}$, $N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta v_i \xrightarrow{d} \tilde{X}_{52}$, $N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta y_{i,-1} \xrightarrow{d} \tilde{X}_{62}$, and $\hat{\rho}_{AB1} - \rho \xrightarrow{d} \tilde{X}'_6 \tilde{\Sigma}_{55}^{-1} \tilde{X}_5 / \tilde{X}'_6 \tilde{\Sigma}_{55}^{-1} \tilde{X}_6$ with $\tilde{X}_5 = (\tilde{X}'_{51} \tilde{X}'_{52})' \sim N(0, \tilde{\Sigma}_{55})$, $\tilde{X}_6 = (\tilde{X}'_{61} \tilde{X}'_{62})' \sim N(\tilde{\mu}_6, \tilde{\Sigma}_{66})$ and $E(\tilde{X}_5 \tilde{X}'_6) = \tilde{\Sigma}_{56} \neq 0$, where $\tilde{\mu}_6$, $\tilde{\Sigma}_{55}$, $\tilde{\Sigma}_{56}$ and $\tilde{\Sigma}_{66}$ are given in the proof; assuming that $\hat{\rho}_{AB}$ exploits $E(Z_i^{AB'} \Delta v_i) = 0$ in lieu of $E(Z_i^I \Delta v_i) = 0$, $\hat{\rho}_{AB} - \rho \xrightarrow{d} \tilde{X}'_{61} W_{11} \tilde{X}_{51} / \tilde{X}'_{61} W_{11} \tilde{X}_{61}$; when $T = 3$, $\tilde{X}_5 = \tilde{X}_{51} = \tilde{X}_1 / \sqrt{\lambda}$ and $\tilde{X}_6 = \tilde{X}_{61} = \tilde{X}_2 / \sqrt{\lambda}$, where \tilde{X}_1 and \tilde{X}_2 are defined in theorem 2.

(b) $N^{-1/2} \sum_{i=1}^N Z_i^{L'} v_i \xrightarrow{d} \tilde{X}_{71}$, $N^{-1} \sum_{i=1}^N Z_i^{L'} y_{i,-1} \xrightarrow{d} \tilde{X}_{81} \sim N(\frac{\sigma^2 \iota}{2}, \frac{\sigma^4}{2\lambda} I)$, $\text{plim}_{N \rightarrow \infty} (\hat{\rho}_{Arbov} - \rho) = 0$ and assuming that $\hat{\rho}_{Arbov}$ exploits $E(\tilde{Z}_i^{II'}[\Delta v_i' v_i']') = 0$ in lieu of $E(Z_i^{II'} v_i) = 0$, $\sqrt{N}(\hat{\rho}_{Arbov} - \rho) \xrightarrow{d} \tilde{X}'_8 W \tilde{X}_7 / \tilde{X}'_8 W \tilde{X}_8$ with $\tilde{X}_7 = (\tilde{X}'_{52} \tilde{X}'_{71})' \sim N(0, \tilde{\Sigma}_{77})$, $\tilde{\Sigma}_{77} = \sigma^4 K^{II} K^{II'}$, $\tilde{X}_8 = (0' \tilde{X}'_{81})'$ and $E(\tilde{X}_7 \tilde{X}'_8) = \tilde{\Sigma}_{78} = 0$; when $T = 3$, $\tilde{X}_7 = \tilde{X}_3$ and $\tilde{X}_8 = \tilde{X}_4$, where \tilde{X}_3 and \tilde{X}_4 are defined in theorem 2.

(c) $\text{plim}_{N \rightarrow \infty} N^{-2} \sum_{i=1}^N (Z_i^{I'} Z_i^I) = \frac{\sigma^2}{2\lambda} I$, $\text{plim}_{N \rightarrow \infty} N^{-3/2} \sum_{i=1}^N (Z_i^{I'} Z_i^D) = 0$ and $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N (Z_i^{D'} Z_i^D) = \sigma^2 I$.

(d) $\text{plim}_{N \rightarrow \infty} W_{N,Arbov1} = \sigma^{-2} I$ and $N^{-1} \sum_{i=1}^N \tilde{Z}_i^{II'} [\widehat{\Delta v}_i' \hat{v}_i']' [\widehat{\Delta v}_i' \hat{v}_i'] \tilde{Z}_i^{II'} \xrightarrow{d} \sigma^4 \tilde{W}_{Arbov2}^{-1} \equiv \sigma^4 K^{II} K^{II'} + (\sigma^4 / 2\lambda) \mathcal{K}^2 \text{diag}(O_{m-(T-2)}, I_{T-2})$.

(e) $\hat{\rho}_{AB1}$ is asymptotically biased downwards.

PROOF

See appendix A.3.

Since \tilde{X}_8 is Gaussian, part (b) of theorem 3 implies that when the data are covariance stationary Arbov estimators have a non-normal local-to-unity asymptotic distribution. Furthermore, the second result in theorem 3 part (d) implies that in this case the conventional

estimator of the optimal weight matrix for the Arbov estimator is inconsistent under local-to-unity asymptotics. As a consequence, the conventional asymptotic standard errors of Arbov estimators are inconsistent as well under these asymptotics.

Note that \tilde{X}_7 and \tilde{X}_8 are uncorrelated Gaussian vectors. Then it is easily seen that the correct asymptotic standard errors of Arbov estimators are given by $E(\tilde{X}'_8 W \tilde{\Sigma}_{77} W \tilde{X}_8 / (\tilde{X}'_8 W \tilde{X}_8)^2)$ and that the (truly) asymptotically optimal weight matrix is proportional to $\tilde{\Sigma}_{77}^{-1}$.⁶ Moreover, although the Arbov estimator that uses $W_{N,Arbov2}(\hat{\rho}_1)$ as weight matrix is still consistent when $T > 3$, its local-to-unity asymptotic distribution is different from that of the truly optimal Arbov estimator. Finally, the local-to-unity asymptotic distributions of Arbov estimators are symmetric since $\tilde{X}_7 \perp \tilde{X}_8$, i.e. \tilde{X}_7 is independent of \tilde{X}_8 .

Under covariance stationarity and the parameter sequence $\rho = 1 - \lambda N^{-1}$, the cross-products of the instruments and the regressors in the AB estimator remain correlated with the cross-products of the instruments and the dependent variables when $N \rightarrow \infty$, thereby causing (explaining) the bias of this estimator. On the other hand, when $T > 3$ the Arbov estimators are asymptotically unbiased under these local-to-unity asymptotics despite the fact that the Arbov moment conditions are weak when the data are covariance stationary.

Note that the set of $m + T - 2$ System moment conditions in $E(Z_i^{S'}[\Delta v'_i \ v'_i]') = 0$ can be rewritten as $E(\tilde{Z}_i^{S'}[\Delta v'_i \ v'_i]') = 0$, where $\tilde{Z}_i^S = \text{diag}(Z_i^{AB}, Z_i^L)$ is a $2(T - 2) \times (m + T - 2)$ matrix. We have the following results for the System estimator:

Theorem 4 *Let the CS model hold, let $T \geq 3$ and let $\rho = 1 - \lambda N^{-1}$ with $\lambda > 0$. Let $\hat{\rho}_1$ be an initial \sqrt{N} -consistent estimator for ρ , i.e. $\sqrt{N}(\hat{\rho}_1 - \rho) \xrightarrow{d} \mathcal{K} \neq 0$. Let W_N be an arbitrary sequence of PDS weight matrices with $\text{plim}_{N \rightarrow \infty} W_N = W$, where W is PDS. Let $\tilde{W}_{SYS^*} = \text{diag}(O_{T-2}, I_m)$, where O_{T-2} is a $(T - 2) \times (T - 2)$ null matrix. Let $W_{N,SYS1b} = [N^{-1} \sum_{i=1}^N (Z_i^{S'} Z_i^S)]^{-1}$. Finally, let $D = E[\text{diag}(\varepsilon_i^2, \dots, \varepsilon_i^{T-1})' \iota \Delta \varepsilon'_i] / \sigma^2$. Then*

(a) $\text{plim}_{N \rightarrow \infty} W_{N,SYS1} = \sigma^{-2} \tilde{W}_{SYS^*}$ and $\text{plim}_{N \rightarrow \infty} W_{N,SYS1b} = \sigma^{-2} \tilde{W}_{SYS^*}$.

(b) $\text{diag}(N^{-1} I_{T-2}, N^{-1/2} I_m) \sum_{i=1}^N \tilde{Z}_i^{S'} [\widehat{\Delta v}'_i \ \widehat{v}'_i] [\widehat{\Delta v}'_i \ \widehat{v}'_i] \tilde{Z}_i^S \text{diag}(N^{-1} I_{T-2}, N^{-1/2} I_m) \xrightarrow{d} \sigma^4 \tilde{W}_{SYS2}^{-1}$, with $\tilde{W}_{SYS2}^{11} \equiv (\frac{1}{2\lambda}) H$, $\tilde{W}_{SYS2}^{21} = (\tilde{W}_{SYS2}^{12})' \equiv -(\frac{1}{2\lambda}) \mathcal{K} K^{II} D$, and $\tilde{W}_{SYS2}^{22} \equiv \tilde{W}_{Arbov2}^{-1}$,

⁶ $E(1/\tilde{X}'_8 \tilde{\Sigma}_{77}^{-1} \tilde{X}_8) \leq E(\tilde{X}'_8 W \tilde{\Sigma}_{77} W \tilde{X}_8 / (\tilde{X}'_8 W \tilde{X}_8)^2)$ for any W .

where \tilde{W}_{Arbov2}^{-1} is defined in theorem 3.

(c) $\sqrt{N}(\hat{\rho}_{SYSk} - \rho) \xrightarrow{d} \tilde{X}'_8 \tilde{\Sigma}_{77}^{-1} \tilde{X}_7 / \tilde{X}'_8 \tilde{\Sigma}_{77}^{-1} \tilde{X}_8$, for $k = 1, 1b$, and $\sqrt{N}(\hat{\rho}_{SYS2} - \rho) \xrightarrow{d} \tilde{X}'_{11} \tilde{W}_{SYS2} \tilde{X}_{10} / \tilde{X}'_{11} \tilde{W}_{SYS2} \tilde{X}_{11}$ with $\tilde{X}_{10} = (\tilde{X}'_{51} \tilde{X}'_{52} \tilde{X}'_{71})'$ and $\tilde{X}_{11} = (0 \ 0 \ \tilde{X}'_{81})'$, where $\tilde{\Sigma}_{77}$, \tilde{X}_7 and \tilde{X}_8 are defined in theorem 3. If $T > 3$, $\text{plim}_{N \rightarrow \infty}(\hat{\rho}_{SYSk} - \rho) = 0$, for $k = 1, 1b, 2$.

(d) $\hat{\rho}_{SYS} - \rho \xrightarrow{d} (\tilde{X}'_{61} W_{11} \tilde{X}_{51} + \tilde{X}'_{81} W_{31} \tilde{X}_{51}) / \tilde{X}'_{12} W \tilde{X}_{12}$ with $\tilde{X}_{12} = (\tilde{X}'_{61} \ 0 \ \tilde{X}'_{81})'$, where \tilde{X}_{51} , \tilde{X}_{61} and \tilde{X}_{81} are defined in theorem 3.

PROOF

See appendix A.4.

Theorem 4 implies that when the data are stationary, a System estimator which uses a weight matrix estimator that has a PDS probability limit, is inconsistent under local-to-unity asymptotics. Moreover, the conventional estimator for the optimal weight matrix for the System estimator, i.e. $W_{N,SYS2}(\hat{\rho}_1)$, and the conventional asymptotic standard errors of System estimators are inconsistent as well under such asymptotics.⁷ Nevertheless, the System estimators that use $W_{N,SYS1}$ and $W_{N,SYS2}(\hat{\rho}_1)$, respectively, as weight matrix are still consistent under such asymptotics although the local-to-unity asymptotic distribution of the two-step System estimator is different from that of the truly optimal System estimator. In particular, the former distribution is asymmetric since $E(\tilde{X}_{51} \tilde{X}'_{81}) \neq 0$. However, the truly optimal System estimator has the same local-to-unity asymptotic distribution as the truly optimal Arbov estimator, which is symmetric. Thus the local-to-unity asymptotic distribution of the truly optimal System estimator is not affected by the AB moment conditions.

The theoretical results above suggest that Wald tests based on the usual two-step System estimator will have incorrect size when the data are stationary and ρ is close to one due to its asymmetric distribution. Bond and Windmeijer (2002) found in a Monte Carlo study for covariance stationary data with $\rho = 0.8$, $T = 6$ and $N = 100$ that the two-step System

⁷Kruiniger (2005) discusses simple estimators for the optimal weight matrices for the Arbov and the System estimators that are consistent under both first-order fixed parameter asymptotics and local asymptotics as well as under a large variety of asymptotic plans for S and N , including fixed S , large N asymptotics, sequential asymptotics with $(S, N \rightarrow \infty)_{seq}$, and diagonal path asymptotics with $(S, N \rightarrow \infty, S/N^d \rightarrow \hat{c} > 0$ where $0 \leq d \leq 1$).

estimator that uses $W_{N,SY S_2}(\widehat{\rho}_1)$ as weight matrix has indeed an asymmetric distribution and that the corresponding Wald tests have incorrect size even when corrected standard errors due to Windmeijer (2005) are used. On the other hand, LM tests which are based on System estimators that use a restricted estimator of the weight matrix that is optimal under the null, e.g. $W_{N,SY S_2}(\rho)$, had rejection frequencies very close to their nominal size. These findings are in agreement with the above theory which implies that LM tests which are based on Arbov or System estimators that use a restricted estimator of the weight matrix that is optimal under the null, e.g. $W_{N,Arbov2}(\rho)$ and $W_{N,SY S_2}(\rho)$, respectively, have a standard normal asymptotic distribution under the null both when first-order fixed parameter asymptotics and when local-to-unity asymptotics are employed owing to the fact that $\tilde{X}_7 \perp \tilde{X}_8$.⁸

3.3 Diagonal path local asymptotics

The results in section 3.2 are largely based on the fact that if $\rho = 1 - \lambda N^{-1}$, then the sequential limits $\lim_{N \rightarrow \infty} \lim_{S \rightarrow \infty} [Var(y_{i,t} - \mu_i)]/N = \sigma^2/2\lambda > 0$, for $t = 1, 2, \dots, T$. Now suppose that $\rho = 1 - \lambda N^{-1}$ and that $N \rightarrow \infty$, $S \rightarrow \infty$ simultaneously with $S/N \rightarrow c > 0$, where c is a constant, then $\lim_{N \rightarrow \infty, S/N \rightarrow c} \rho^{2(1+S)} = \exp(-2\lambda c) < 1$, since $\lim_{p \rightarrow \infty} (1+x/p)^p = \exp(x)$. It follows that $\lim_{N \rightarrow \infty, S/N \rightarrow c} [Var(y_{i,t} - \mu_i)]/N = \sigma^2[1 - \exp(-2\lambda c)]/2\lambda > 0$. This result suggests that when S is of the same order of magnitude as N and the data are persistent, then the AB and Arbov moment conditions are still weak. Moreover, in this case one can still obtain approximations to the distributions of the AB, Arbov and SYS estimators by using local asymptotics.

It turns out that we can derive diagonal path local asymptotic results for $N, S \rightarrow \infty$ with $(S/N^d) \rightarrow \hat{c} > 0$ for any $d \geq 0$. Let $q(\lambda, c) = 1 - \exp(-2\lambda c)$. For brevity we only state the following diagonal path local asymptotic results for the AB and Arbov estimators and $T = 3$:

Theorem 5 *Let $T = 3$ and let $\rho = 1 - \lambda N^{-g}$ with $\lambda > 0$ and $0 < g \leq 1$. Let $0 \leq d \leq g$ and let $\Delta_{gd} = \frac{1}{2}(1 - d) - (g - d)$. Moreover, let $N \rightarrow \infty$, $S \rightarrow \infty$ simultaneously with*

⁸Conditional on \tilde{X}_8 , these LM test-statistics have a standard normal asymptotic distribution.

As the latter distribution does not depend on \tilde{X}_8 , it follows that the unconditional asymptotic distribution of these LM test-statistics is equal to the standard normal distribution as well.

$S/N \rightarrow c \geq 0$, $S/N^g \rightarrow \bar{c} \geq 0$ and $S/N^d \rightarrow \hat{c} > 0$. Finally, let $\hat{q}(\lambda, \hat{c}) = 2\lambda(S + 1)$ if $d = 0$; $\hat{q}(\lambda, \hat{c}) = 2\lambda\hat{c}$ if $0 < d < g$; and $\hat{q}(\lambda, \hat{c}) = q(\lambda, \hat{c})$ if $d = g$. Then

(a) $N^{\Delta_{gd}}(\hat{\rho}_{AB} - \rho) \xrightarrow{d} N(0, 2\sigma^2\sigma_y^2/(\lambda^2(\sigma_{\mu y} - \sigma_y^2)^2))$ if $\Delta_{gd} > 0$ and $d = 0$,

$N^{\Delta_{gd}}(\hat{\rho}_{AB} - \rho) \xrightarrow{d} N(0, 4/(\lambda\hat{q}(\lambda, \hat{c})))$ if $\Delta_{gd} > 0$ and $d > 0$,

$\hat{\rho}_{AB} \xrightarrow{d} \rho + \frac{\check{X}_1}{\check{X}_2}$ if $\Delta_{gd} = 0$ and $d = 0$, where \check{X}_1 and \check{X}_2 are defined in theorem 1,

$\hat{\rho}_{AB} \xrightarrow{d} \rho + \frac{\hat{X}_1}{\hat{X}_2}$, with $\begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \end{bmatrix} \sim N \left[-\frac{\sqrt{\lambda\hat{q}(\lambda, \hat{c})}}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right]$ if $\Delta_{gd} = 0$ and $d > 0$,

$\hat{\rho}_{AB} \xrightarrow{d} \rho + \frac{\hat{X}_1}{\hat{X}_2}$, with $\begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right]$ if $\Delta_{gd} < 0$,

(b) $\sqrt{N}(\hat{\rho}_{Arbov} - \rho) \xrightarrow{d} \frac{\check{X}_3}{\check{X}_4}$, with $\begin{bmatrix} \hat{X}_3 \\ \hat{X}_4 \end{bmatrix} \sim N \left[\frac{1}{2} \begin{pmatrix} 0 \\ [2 - q(\lambda, \bar{c})] \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \frac{q(\lambda, \bar{c})}{2\lambda} \end{pmatrix} \right]$.

PROOF

See appendix A.5.

Note that $\Delta_{gd} = \frac{1}{2}(1 + d) - g$ and that $\Delta_{gd} < 1/2$. When the values of d and g are such that $\Delta_{gd} = 0$, ρ is weakly identified by the AB moment condition(s). When $\Delta_{gd} > 0$, ρ is nearly weakly identified and when $\Delta_{gd} < 0$, ρ is nearly non-identified by the AB moment condition(s).

When $0 \leq d \leq 1$ and $\Delta_{gd} = 0$, there are two extreme cases: if $d = 0$, then one obtains the local-to-nonidentification asymptotic distribution of $\hat{\rho}_{AB}$ for $g = 1/2$. This case corresponds to local-to-zero asymptotics, see theorem 1. On the other hand, if $d = 1$, then one obtains the local-to-nonidentification asymptotic distribution of $\hat{\rho}_{AB}$ for $g = 1$. When $\Delta_{gd} = 0$ and the value of d increases from 0 to 1, both the signal, $E(y_{i,1}\Delta y_{i,2}) = O(N^{d-g})$, and the noise of the AB moment function, $[Var(y_{i,1}\Delta y_{i,2})]^{1/2} = O(N^{\frac{1}{2}d})$, become stronger.

The diagonal path local (-to-nonidentification) asymptotic results of theorem 5 are very similar to the sequential asymptotic results of theorem 2, which were obtained for $g = 1$. When $d = 1$ and $\Delta_{gd} = 0$ (so that $g = 1$), $\hat{q}(\lambda, \hat{c}) = q(\lambda, c)$ and the sequential local-to-unity asymptotic distributions of $\hat{\rho}_{AB}$ and $\hat{\rho}_{Arbov}$ as $(S, N \rightarrow \infty)_{seq}$ could be obtained from

theorem 5 by letting $c \rightarrow \infty$ so that $\hat{q}(\lambda, \hat{c}) = q(\lambda, c) = q(\lambda, \bar{c}) = 1$. Indeed, when $d > 1$ the non-normal local (-to-nonidentification) asymptotic distributions of $\hat{\rho}_{AB}$ and $\hat{\rho}_{Arbov}$ are also obtained for $g = 1$ and they can be shown to be equal to the sequential local-to-unity asymptotic distributions of $\hat{\rho}_{AB}$ and $\hat{\rho}_{Arbov}$ as $(S, N \rightarrow \infty)_{seq}$. On the other hand, the first-order large N fixed S asymptotic distribution of $\hat{\rho}_{Arbov}$ for $\rho = 1$ can be obtained from theorem 5 by assuming that $\bar{c} = 0$, i.e. $d < g \leq 1$, so that $q(\lambda, \bar{c}) = q(\lambda, c) = 0$. Finally, note that $\frac{\partial \hat{q}(\lambda, \hat{c})}{\partial \hat{c}} > 0$.

When $0 \leq d \leq 1$ (and $\Delta_{gd} = 0$) one can easily obtain diagonal path local(-to-nonidentification) asymptotic counterparts of theorems 3 and 4 by adjusting some rates of convergence and by adjusting the formulae for the presence of $\hat{q}(\lambda, \hat{c})$, $q(\lambda, \bar{c})$ and $q(\lambda, c)$:⁹ in fact, the results in theorems 3 and 4 remain valid under diagonal path local (-to-nonidentification) asymptotics, apart from the fact that $\tilde{X}_5 (= \hat{X}_5) = \hat{X}_1 \sqrt{\hat{q}(\lambda, \hat{c})/\lambda}$ and $\tilde{X}_6 (= \hat{X}_6) = \hat{X}_2 \sqrt{\hat{q}(\lambda, \hat{c})/\lambda}$ when $T = 3$, where \hat{X}_1 and \hat{X}_2 are defined in theorem 5, $\tilde{X}_{81} \sim N(\frac{\sigma^2 \epsilon}{2} [2 - q(\lambda, \bar{c})], \frac{\sigma^4 q(\lambda, c)}{2\lambda} I)$, $\text{plim}_{N \rightarrow \infty} N^{-1-d} \sum_{i=1}^N (Z_i' Z_i) = \frac{\sigma^2 \hat{q}(\lambda, \hat{c}) + 1 \{d=0\} E(y_{i,-S}^2)}{2\lambda} I$, and some straightforward adjustments of some rates of convergence and the formulae for \tilde{W}_{Arbov2}^{-1} and $\tilde{W}_{SY^2}^{-1}$. When $d \geq 1$ and $g = 1$ the estimators of the optimal weight matrices for the Arbov estimator and the System estimator and the asymptotic standard errors of these estimators are inconsistent, and the local asymptotic distribution of the two-step ‘optimal’ System estimator is skewed.

To see that the above asymptotic results could be empirically relevant, consider the following example: let $\rho = 0.95$, $N = 100$ and $S = 10$. Choose $d = g = 1$. Then $\lambda = 5$, $c = 1/10$, and $q(\lambda, c) = 1 - \exp(-1) = 0.632$, while $1 - \rho^{2(1+S)} = 0.676$.

We now consider the diagonal path asymptotic distributions of the AB and Arbov estimators for $\rho = 1$:

Theorem 6 *Let $T = 3$ and let $\rho = 1$. In addition, let $N \rightarrow \infty$, $S \rightarrow \infty$ simultaneously with $S/N^d \rightarrow \hat{c} > 0$ and $S/N^{\bar{d}} \rightarrow c \geq 0$ where $d \geq 0$ and $\bar{d} = \max(d, 1)$. Finally, let $1\{d \leq 1\} = 1$ if $d \leq 1$ and let $1\{d \leq 1\} = 0$ if $d > 1$. Then*

$$(a) \hat{\rho}_{AB} - \rho \xrightarrow{d} \frac{\hat{X}_1}{\hat{X}_2}, \quad \text{where} \quad \begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right],$$

⁹Note that when $d > 1$, no adjustments are required.

$$(b) N^{\frac{1}{2}d}(\hat{\rho}_{Arbov} - \rho) \xrightarrow{d} \frac{\hat{X}_3}{\hat{X}_4}, \quad \text{where} \quad \begin{bmatrix} \hat{X}_3 \\ \hat{X}_4 \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ 1\{d \leq 1\} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix} \right].$$

PROOF

See appendix A.6.

The results in theorem 6 can easily be extended to $T > 3$ and to the System estimator.

One obtains the first-order large N fixed S asymptotic distributions of the estimators for $\rho = 1$ by taking $c = 0$. However, in general c is unknown and could well be strictly positive in which case the Arbov estimator has a non-normal asymptotic distribution for $\rho = 1$.

Observe that $\lim_{\lambda \downarrow 0} \hat{q}(\lambda, \hat{c}) = 0$ and $\lim_{\lambda \downarrow 0} [q(\lambda, c)/2\lambda] = \lim_{\lambda \downarrow 0} \{[1 - \exp(-2\lambda c)]/2\lambda\} = \lim_{\lambda \downarrow 0} \{2c \exp(-2\lambda c)/2\} = c$. From these observations and the results in theorems 5 and 6 we conclude that the diagonal path local asymptotic distributions of the Arbov and SYS estimators are continuous at $\lambda = 0$ ($\rho = 1$) provided that $0 \leq d \leq 1$. A similar continuity result holds for the AB estimator for any $d \geq 0$.

4 LM panel unit root tests

In this section we propose two LM-type panel unit root (UR) test statistics that are based on an Arbov estimator and a System estimator which use a weight matrix that is optimal under the null, and on restricted conventional estimators of their first-order fixed parameter asymptotic standard errors. The critical values for these LM tests can be taken from the standard normal distribution irrespective of the assumptions made regarding the initial observations.

Let $\hat{\rho}_{Arbov2,R}$ and $\hat{\rho}_{SYS2,R}$ be GMM estimators that use $W_{N,Arbov2}(1)$ and $W_{N,SYS2}(1)$ as weight matrix, respectively. Let $SE(\hat{\rho}_{Arbov2,R}) = \{N^{-1} \sum_{i=1}^N (y'_{i,-1} Z_i^{II}) [W_{N,Arbov2}(1)] \times \sum_{i=1}^N (Z_i^{III} y_{i,-1})\}^{-1/2}$ and $SE(\hat{\rho}_{SYS2,R}) = \{N^{-1} \sum_{i=1}^N [(\Delta y'_{i,-1} \quad y'_{i,-1}) Z_i^S] [W_{N,SYS2}(1)] \times \sum_{i=1}^N [Z_i^{S'} (\Delta y'_{i,-1} \quad y'_{i,-1})']\}^{-1/2}$. Then we have the following results:

Theorem 7 *Let $T \geq 3$.*

(a) *If $\rho = 1$ and if $N \rightarrow \infty$, $S \rightarrow \infty$ simultaneously with $S/N^d \rightarrow \hat{c} > 0$ for some $d \geq 0$, then $(\hat{\rho}_{Arbov2,R} - 1)/SE(\hat{\rho}_{Arbov2,R}) \xrightarrow{d} N(0, 1)$ and $(\hat{\rho}_{SYS2,R} - 1)/SE(\hat{\rho}_{SYS2,R}) \xrightarrow{d} N(0, 1)$.*

(b) *If $\rho = 1 - \lambda N^{-1/2}$ with $\lambda > 0$ and if $S \rightarrow \infty$, $N \rightarrow \infty$ sequentially, then $(\hat{\rho}_{Arbov2,R} - 1)/SE(\hat{\rho}_{Arbov2,R}) \xrightarrow{d} N(-\frac{\lambda}{2}\sqrt{m}, 1)$ and $(\hat{\rho}_{SYS2,R} - 1)/SE(\hat{\rho}_{SYS2,R}) \xrightarrow{d} N(-\frac{\lambda}{2}\sqrt{m}, 1)$.*

(c) If $\rho = 1 - \lambda N^{-1/2}$ with $\lambda \geq 0$ and if $N \rightarrow \infty$, $S \rightarrow \infty$ simultaneously with $S/N^{1/2} \rightarrow c \geq 0$ and $S/N^d \rightarrow \hat{c} > 0$ where $d \geq 0$, then $(\hat{\rho}_{Arbov2,R} - 1)/SE(\hat{\rho}_{Arbov2,R}) \xrightarrow{d} N(-\lambda(1 - q(\lambda, c)/2)\sqrt{m}, 1)$ and $(\hat{\rho}_{SYS2,R} - 1)/SE(\hat{\rho}_{SYS2,R}) \xrightarrow{d} N(-\lambda(1 - q(\lambda, c)/2)\sqrt{m}, 1)$.

PROOF

See appendix A.7. The results in theorem 7(c) corresponding to $c = 0$ are also valid under large N fixed S asymptotics. Madsen (2003) has derived some related local power results.

Note that the local power is the lowest when the data are covariance stationary or $d > 1/2$.

Any Arbov estimator can be used to construct an LM-type panel UR test but choosing the optimal Arbov estimator $\hat{\rho}_{Arbov2,R}$ yields a test statistic with the highest local power within this class of LM tests. In particular, in the proof of theorem 7 it is shown that the weight matrix used by $\hat{\rho}_{Arbov2,R}$ is not only optimal under the null but also optimal under local alternatives given by $\rho = 1 - \lambda N^{-1/2}$ with $\lambda > 0$, irrespective of the asymptotic plan for S and N . On the other hand, not every System estimator is suitable for the construction of an LM-type panel UR test that has correct size for any value of d and nontrivial power against any local alternative. However, the optimal System estimator $\hat{\rho}_{SYS2,R}$ is \sqrt{N} -consistent under both the null and local alternatives and yields an LM test that has correct size and the same local power properties as the LM test which is based on the optimal Arbov estimator $\hat{\rho}_{Arbov2,R}$, irrespective of the asymptotic plan for S and N .

In the last decade various other panel UR tests have been proposed. For instance, Breitung and Meyer (1994) proposed a test-statistic which is based on an OLS estimator for ρ in a model for deviations from the initial observations. Harris and Tzavalis (1999) discussed an LM-type panel UR test which is based on the bias-corrected LSDV estimator for ρ . Finally, Kruginer (2004) discussed a Wald-type panel UR test which is based on the First Difference MLE for ρ in the covariance stationary panel AR(1)/UR model.

4.1 Monte Carlo results

In this section we compare the finite sample performance of our GMM based panel UR tests with three other panel UR tests, namely the Wald test which is based on the FDMLE, the LM test due to Harris and Tzavalis (1999) which is based on the bias corrected LSDV estimator, and a Wald test which is based on the Fixed Effects MLE for r in the panel

AR(1) model (cf Hsiao et al., 2002, and Kruiniger, 2001 and 2004).^{10 11} We only consider size-adjusted versions of the latter two tests. The implementation of the FEMLE based UR test is further discussed in Kruiniger (2004).

In most simulation experiments the errors have been drawn from normal distributions: $\varepsilon_{i,t} \sim N(0, 1)$ and $\mu_i \sim N(0, 1)$. To assess how assumptions with respect to $y_{i,1} - \mu_i$, $i = 1, \dots, N$, affect the power of the tests, we have conducted four different kinds of experiments: in one set the initial observations are drawn from stationary distributions, i.e. $(y_{i,1} - \mu_i) | \mu_i \sim N(0, 1/(1 - r^2))$, while in the other three sets the initial observations are non-stationary. The three non-stationary cases considered are: (1) $y_{i,1} - \mu_i = 0$; (2) $(y_{i,1} - \mu_i) | \mu_i \sim N(0, 2/(1 - r^2))$; and (3) $(y_{i,1} - 2\mu_i) | \mu_i \sim N(0, r^2/(1 - r^2))$. Note that in all situations $E(y_{i,t} - y_{i,t-1}) = 0$ as is the case under the null hypothesis. In both case (1) and case (2) the variance of $y_{i,1} - \mu_i$ is different from the variance under stationarity, while in case (3) nonstationarity is due to the fact that $E[\mu_i(y_{i,1} - \mu_i)] \neq 0$. Case (1) corresponds to small S . We have also considered experiments with $\sigma_\mu^2 = 0$, $\sigma_\mu^2 = 100$ or $\varepsilon_{i,t} \sim (\chi^2(1) - 1)/\sqrt{2}$. Note that the size and the power of test-statistics which only exploit data in differences are not affected by changes in σ_μ^2 . Finally, in the simulation experiments we have varied the dimensions of the panel data sets as well: $(N, T) = (100, 10)$, $(100, 6)$, or $(500, 6)$. All simulation results are based on 5,000 replications and the (nominal) level of the tests is either 2.5% or 5%.

Tables 1-7 report the simulation results on the empirical size and power of the panel UR tests that were mentioned above. Tables 2, 5, 6 and 7 report results on power against stationary alternatives, whereas tables 1, 3 and 4 report results on power against non-stationary alternatives. In the tables ‘W’ denotes the Wald version of a test, ‘LM’ stands for the LM version of a test, whilst ‘SA’ indicates that the test has been size adjusted. When

¹⁰Note that the FEMLE is derived under the imposed assumption that $\text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i,1} - \mu_i)^2 < \infty$ and is different from the LSDV estimator.

¹¹Bond et al. (2005) have also compared various panel unit root tests by calculating their asymptotic local power and by conducting simulation experiments. The only GMM based test they consider is a Wald test based on the ‘optimal’ System estimator. As expected this test has poor size and power properties. They also find that the Breitung-Meyer test and the test based on the FDMLE have very similar size and power properties.

the (nominal) level of a test is 2.5% this is indicated by a , otherwise the level of a test is 5%. Inspection of the results in tables 1-7 leads to the following conclusions with respect to the GMM based UR tests:

1. In most cases considered the GMM based tests have correct size. However non-normality of the errors affects the size of the tests.
2. The power of the test based on the System estimator is greater than or equal to the power of the test based on the Arbov estimator. However, in many cases the power of both tests is roughly the same and equal to the power of the FDMLE.
3. When the variance of the $y_{i,1} - \mu_i$ is larger than the value implied by covariance stationarity, the power of the test based on the System estimator is greater than the power of the tests based on the Arbov estimator and the FDMLE.
4. The power of the GMM based tests decreases with an increase of the value of σ_μ^2 .
5. In the cases considered the GMM based tests have greater power than either the test that is based on the LSDV estimator or the test that is based on the FEMLE.

5 Concluding remarks

In this paper we considered GMM based estimation and inference for the panel AR(1) model when the data are persistent and the time dimension of the panel is fixed. We derived local asymptotic approximations to the finite sample distributions of the AB, Arbov and System estimators, respectively, under a variety of distributional assumptions about the initial observations. Among other things we found that the nature of the weak instruments problem of the Arellano-Bond estimator depends on the distributional properties of the initial observations. Moreover, when $\rho = 1 - \lambda/N$ and when either the data are covariance stationary or both S and N grow large with $S/N^d \rightarrow c > 0$ and $d \geq 1$, then both the Arbov and the two-step ‘optimal’ System estimator have non-normal local asymptotic distributions, and the estimators of the optimal weight matrices for the Arbov estimator and the System estimator and their asymptotic standard errors are no longer consistent. We also argued that in these cases one should use LM tests and not Wald tests. Two LM-type panel unit root tests that we proposed were found to have good size and power properties.

T	N	test	size, $S = -1$	size, $S = 49$	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
6	100	ARBOV2-LM ^a	0.035	0.030	0.308	0.799	0.999
		ARBOV2-LM	0.058	0.056	0.416	0.872	0.999
		SYS2-LM ^a	0.029	0.030	0.281	0.759	0.998
		SYS2-LM	0.050	0.054	0.378	0.842	0.999
		FDML-W	0.056	0.056	0.457	0.914	1.000
		FEML-W-SA	0.049	0.049	0.138	0.182	0.290
		LSDV-LM-SA	0.050	0.050	0.308	0.732	0.994

Table 1: power against “non-stationary” alternatives with $y_{i,1} = \mu_i$.

T	N	test	size, $S = -1$	size, $S = 49$	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
10	100	ARBOV2-LM ^a	0.030	0.029	0.261	0.663	0.987
		ARBOV2-LM	0.059	0.054	0.370	0.756	0.993
		SYS2-LM ^a	0.024	0.028	0.209	0.592	0.981
		SYS2-LM	0.049	0.047	0.306	0.700	0.989
6	100	ARBOV2-LM ^a	0.035	0.030	0.145	0.339	0.844
		ARBOV2-LM	0.058	0.056	0.217	0.459	0.901
		SYS2-LM ^a	0.029	0.030	0.139	0.341	0.856
		SYS2-LM	0.050	0.054	0.209	0.453	0.916
		FDML-W	0.056	0.056	0.200	0.466	0.925
		FEML-W-SA	0.049	0.049	0.118	0.162	0.319
		LSDV-LM-SA	0.050	0.050	0.153	0.327	0.812
6	500	ARBOV2-LM ^a	0.029	0.029	0.446	0.939	1.000
		ARBOV2-LM	0.055	0.053	0.553	0.968	1.000
		SYS2-LM ^a	0.027	0.028	0.446	0.958	1.000
		SYS2-LM	0.050	0.052	0.567	0.979	1.000

Table 2: power against stationary alternatives.

T	N	test	size, $S = -1$	size, $S = 49$	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
6	100	ARBOV2-LM ^a	0.035	0.030	0.096	0.138	0.359
		ARBOV2-LM	0.058	0.056	0.157	0.217	0.473
		SYS2-LM ^a	0.029	0.030	0.089	0.189	0.609
		SYS2-LM	0.050	0.054	0.151	0.278	0.721
		FDML-W	0.056	0.056	0.068	0.116	0.351
		FEML-W-SA	0.049	0.049	0.098	0.170	0.396
		LSDV-LM-SA	0.050	0.050	0.060	0.101	0.378

Table 3: power against “non-stationary” alternatives with $(y_{i,1} - \mu_i) | \mu_i \sim N(0, 2/(1 - \rho^2))$.

T	N	test	size, $S = -1$	size, $S = 49$	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
6	100	ARBOV2-LM ^a	0.035	0.030	0.138	0.351	0.799
		ARBOV2-LM	0.058	0.056	0.216	0.459	0.870
		SYS2-LM ^a	0.029	0.030	0.133	0.355	0.826
		SYS2-LM	0.050	0.054	0.205	0.461	0.893
		FDML-W	0.056	0.056	0.205	0.474	0.927
		FEML-W-SA	0.049	0.049	0.118	0.166	0.325
		LSDV-LM-SA	0.050	0.050	0.143	0.335	0.811

Table 4: power against “non-stationary” alternatives with $(y_{i,1} - 2\mu_i)|\mu_i \sim N(0, \rho^2/(1 - \rho^2))$ and $\mu_i \sim N(0, 1)$.

T	N	test	size, $S = -1$	size, $S = 49$	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
6	100	ARBOV2-LM ^a	0.035	0.030	0.155	0.355	0.846
		ARBOV2-LM	0.058	0.056	0.228	0.470	0.906
		SYS2-LM ^a	0.029	0.030	0.143	0.370	0.881
		SYS2-LM	0.050	0.054	0.215	0.477	0.930

Table 5: power against stationary alternatives, $\sigma^2 = 1$ and $\sigma_\mu^2 = 0$.

T	N	test	size, $S = -1$	size, $S = 49$	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
6	100	ARBOV2-LM ^a	0.035	0.030	0.085	0.161	0.376
		ARBOV2-LM	0.058	0.056	0.138	0.236	0.459
		SYS2-LM ^a	0.029	0.030	0.077	0.147	0.351
		SYS2-LM	0.050	0.054	0.127	0.215	0.435

Table 6: power against stationary alternatives, $\sigma^2 = 1$ and $\sigma_\mu^2 = 100$.

T	N	test	size, $S = -1$	size, $S = 49$	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$
6	100	ARBOV2-LM ^a	0.052	0.047	0.163	0.311	0.619
		ARBOV2-LM	0.100	0.091	0.256	0.420	0.732
		SYS2-LM ^a	0.036	0.039	0.141	0.328	0.743
		SYS2-LM	0.070	0.075	0.228	0.447	0.831

Table 7: power against stationary alternatives with $(y_{i,1} - \mu_i)|\mu_i \sim [\chi^2(1) - 1]/2^{1/2}$ and $\mu_i \sim N(0, 1)$.

A Proofs of the results

A.1 Proof of theorem 1

Part a1): When $T = 3$, $\hat{\rho}_{AB} = \frac{N^{-1} \sum_{i=1}^N y_{i,1} \Delta y_{i,3}}{N^{-1} \sum_{i=1}^N y_{i,1} \Delta y_{i,2}} = \rho + \frac{N^{-1} \sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3}}{N^{-1} \sum_{i=1}^N y_{i,1} \Delta y_{i,2}}$. Note that

$$E(y_{i,1} \Delta y_{i,2} | y_{i,1}, \mu_i) = (\rho - 1) y_{i,1} (y_{i,1} - \mu_i), \quad Var(y_{i,1} \Delta y_{i,2} | y_{i,1}, \mu_i) = Var(y_{i,1} \varepsilon_{i,2} | y_{i,1}, \mu_i) = \sigma^2 y_{i,1}^2, \\ E(y_{i,1} \Delta \varepsilon_{i,3}) = 0, \quad Var(y_{i,1} \Delta \varepsilon_{i,3}) = 2\sigma^2 \sigma_y^2 \text{ and } Cov(y_{i,1} \Delta \varepsilon_{i,3}, y_{i,1} \Delta y_{i,2}) = -\sigma^2 \sigma_y^2.$$

Let us define $X_1 = N^{-1/2} \sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3}$ and $X_2 = N^{-1/2} \sum_{i=1}^N y_{i,1} \Delta y_{i,2}$. Then we obtain for the parameter sequence $\rho = 1 - \lambda N^{-1/2}$ that $X_1 \xrightarrow{d} N(0, 2\sigma^2 \sigma_y^2)$ and $X_2 \xrightarrow{d} N(-\lambda(\sigma_y^2 - \sigma_{\mu y}), \sigma^2 \sigma_y^2)$. Moreover, it is easily verified that $Cov(X_1, X_2) = -\sigma^2 \sigma_y^2$.

It follows that $\hat{\rho}_{AB} \xrightarrow{d} \rho + \frac{\check{X}_1}{\check{X}_2}$, where $\begin{bmatrix} \check{X}_1 \\ \check{X}_2 \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ \lambda(\sigma_{\mu y} - \sigma_y^2) \end{pmatrix}, \sigma^2 \sigma_y^2 \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right]$.

Part a2): Similar to part a1). For instance, let $X_{5,p} = N^{-1/2} \sum_{i=1}^N y_{i,k} \Delta \varepsilon_{i,l}$ and $X_{6,p} = N^{-1/2} \sum_{i=1}^N y_{i,k} \Delta y_{i,l-1}$ with $k \leq l - 2$. Note that $E(y_{i,k} \Delta y_{i,l-1}) = (\rho - 1) E[y_{i,k} (y_{i,l-2} - \mu_i)] = (\rho - 1) \rho^{l-2-k} E[(y_{i,k} - \mu_i + \mu_i)(y_{i,k} - \mu_i)]$. Moreover, let $\tilde{\sigma}_{y,k}^2 = \sigma_y^2 + (k-1)\sigma^2$. Then it is easily

seen that for $\rho = 1 - \lambda N^{-1/2}$, $\begin{bmatrix} X_{5,p} \\ X_{6,p} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \check{X}_{5,p} \\ \check{X}_{6,p} \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ \check{\mu}_{6,p} \end{pmatrix}, \sigma^2 \tilde{\sigma}_{y,k}^2 \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right]$,

where $\check{\mu}_{6,p} = \lambda(\sigma_{\mu y} - \tilde{\sigma}_{y,k}^2)$. Finally, let $s \leq t - 2$. Then the (p, q) -th elements of $\check{\Sigma}_{55} = Var(\check{X}_5) = E(\check{X}_5 \check{X}_5')$, $\check{\Sigma}_{66} = Var(\check{X}_6)$, and $\check{\Sigma}_{56} = Cov(\check{X}_5, \check{X}_6) = E(\check{X}_5 \check{X}_6')$ are given by, respectively:

$Cov(\check{X}_{5,p}, \check{X}_{5,q}) = \lim_{N \rightarrow \infty} Cov(y_{i,k} \Delta \varepsilon_{i,l}, y_{i,s} \Delta \varepsilon_{i,t}) = \lim_{N \rightarrow \infty} E(y_{i,k} y_{i,s} \Delta \varepsilon_{i,l} \Delta \varepsilon_{i,t})$, where

$$\lim_{N \rightarrow \infty} E(y_{i,k} y_{i,s} \Delta \varepsilon_{i,l} \Delta \varepsilon_{i,t}) = 2\sigma^2 (\sigma_y^2 + \sigma^2 [-1 + \min(k, s)]) \text{ if } l = t,$$

$$\lim_{N \rightarrow \infty} E(y_{i,k} y_{i,s} \Delta \varepsilon_{i,l} \Delta \varepsilon_{i,t}) = -\sigma^2 (\sigma_y^2 + \sigma^2 [-1 + \min(k, s)]) \text{ if } |l - t| = 1, \text{ and}$$

$$\lim_{N \rightarrow \infty} E(y_{i,k} y_{i,s} \Delta \varepsilon_{i,l} \Delta \varepsilon_{i,t}) = 0 \text{ if } |l - t| > 1;$$

$Cov(\check{X}_{6,p}, \check{X}_{6,q}) = \lim_{N \rightarrow \infty} Cov(y_{i,k} \Delta y_{i,l-1}, y_{i,s} \Delta y_{i,t-1}) = \lim_{N \rightarrow \infty} E(y_{i,k} y_{i,s} \varepsilon_{i,l-1} \varepsilon_{i,t-1})$, where

$$\lim_{N \rightarrow \infty} E(y_{i,k} y_{i,s} \varepsilon_{i,l-1} \varepsilon_{i,t-1}) = \sigma^2 (\sigma_y^2 + \sigma^2 [-1 + \min(k, s)]) \text{ if } l = t, \text{ and}$$

$$\lim_{N \rightarrow \infty} E(y_{i,k} y_{i,s} \varepsilon_{i,l-1} \varepsilon_{i,t-1}) = 0 \text{ if } |l - t| > 0; \text{ and}$$

$Cov(\check{X}_{5,p}, \check{X}_{6,q}) = \lim_{N \rightarrow \infty} Cov(y_{i,k} \Delta \varepsilon_{i,l}, y_{i,s} \Delta y_{i,t-1}) = \lim_{N \rightarrow \infty} E(y_{i,k} y_{i,s} \varepsilon_{i,t-1} \Delta \varepsilon_{i,l})$, where

$$\lim_{N \rightarrow \infty} E(y_{i,k} y_{i,s} \varepsilon_{i,t-1} \Delta \varepsilon_{i,l}) = \sigma^2 (\sigma_y^2 + \sigma^2 [-1 + \min(k, s)]) \text{ if } l = t - 1,$$

$$\lim_{N \rightarrow \infty} E(y_{i,k} y_{i,s} \varepsilon_{i,t-1} \Delta \varepsilon_{i,l}) = -\sigma^2 (\sigma_y^2 + \sigma^2 [-1 + \min(k, s)]) \text{ if } l = t, \text{ and}$$

$$\lim_{N \rightarrow \infty} E(y_{i,k} y_{i,s} \varepsilon_{i,t-1} \Delta \varepsilon_{i,l}) = 0 \text{ if } l \neq t - 1 \text{ and } l \neq t.$$

Recall that $Z'_i = K^{AB} Z_i^{AB'}$, where $Z_i^{AB} = [Z_i^I \ Z_i^D]$. The above results imply that

$$\begin{aligned} & \begin{bmatrix} N^{-1/2} \sum_{i=1}^N Z_i^{I'} \Delta v_i \\ N^{-1/2} \sum_{i=1}^N Z_i^{I'} \Delta y_{i,-1} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \check{X}_{51} \\ \check{X}_{61} \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ \lambda(\sigma_{\mu y} - \sigma_y^2) \end{pmatrix}, \sigma^2 \sigma_y^2 \begin{pmatrix} H & -C' \\ -C & I \end{pmatrix} \right], \text{ and} \\ & \begin{bmatrix} N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta v_i \\ N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta y_{i,-1} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \check{X}_{52} \\ \check{X}_{62} \end{bmatrix} \sim \\ & N \left[\begin{pmatrix} 0 \\ -\lambda \sigma^2 \iota \end{pmatrix}, \sigma^4 \begin{pmatrix} \text{diag}(H_{T-3}, H_{T-4}, \dots, H_1) & -\text{diag}(C_{T-3}, C_{T-4}, \dots, C_1)' \\ -\text{diag}(C_{T-3}, C_{T-4}, \dots, C_1) & I \end{pmatrix} \right]. \end{aligned}$$

Moreover, they imply that $\check{\Sigma}_{55,12} \equiv \text{Cov}(\check{X}_{51}, \check{X}_{52}) = 0$, and also that $\check{\Sigma}_{56,12} = \check{\Sigma}_{56,21} = \check{\Sigma}_{66,12} = 0$. Finally note that $\check{X}_5 = K^{AB} (\check{X}'_{51} \ \check{X}'_{52})'$ and $\check{X}_6 = K^{AB} (\check{X}'_{61} \ \check{X}'_{62})'$.

Part b): When S is fixed, $\hat{\rho}_{Arbov}$ does not suffer from a weak instruments problem. Let $\varepsilon_i^t = [\varepsilon_{i,2} \ \dots \ \varepsilon_{i,t}]$. When $\rho = 1$, $v_i = \varepsilon_i$, $\Delta y_i^t = \varepsilon_i^t$ and $Z_i^{II} = \text{diag}(\varepsilon_i^2, \dots, \varepsilon_i^{T-1})$. It follows that for $\rho = 1 - \lambda N^{-1/2}$, $N^{-1/2} \sum_{i=1}^N Z_i^{II'} v_i \xrightarrow{d} N(0, \sigma^4 I_m)$, $N^{-1} \sum_{i=1}^N Z_i^{II'} y_{i,-1} \xrightarrow{q.m.} \sigma^2 \iota_m$ and hence $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N Z_i^{II'} y_{i,-1} = \sigma^2 \iota_m$. The local-to-unity asymptotic distribution of $\hat{\rho}_{Arbov}$ is equal to the first-order fixed-parameter asymptotic distribution of $\hat{\rho}_{Arbov}$ for $\rho = 1$.

Part c): It is easy to verify that the sequences of sample averages in c), d) and e) converge in probability to the corresponding population means for $\rho = 1$. Therefore we will only prove the claims made with respect to those limits. We first show that $\lim_{N \rightarrow \infty} E(Z_i' H Z_i)$ is PDS. Recall that $Z'_i = K^{AB} Z_i^{AB'}$, where K^{AB} is a nonsingular constant matrix and $Z_i^{AB} = [y_{i,1} I_{T-2} \ d_1 \Delta y_{i,2} \ d_2 \Delta y_{i,3} \ \dots \ d_{T-3} \Delta y_{i,T-2}]$. We have $\lim_{N \rightarrow \infty} E(Z_i^{AB'} H Z_i^{AB}) = \text{diag}(\sigma_y^2 H_{T-2}, \sigma^2 H_{T-3}, \sigma^2 H_{T-4}, \dots, \sigma^2 H_1)$ is PDS since H_t is PDS for all $t \leq T-2$. It follows that $\lim_{N \rightarrow \infty} E(Z_i' H Z_i) = \lim_{N \rightarrow \infty} K^{AB} E(Z_i^{AB'} H Z_i^{AB}) K^{AB'}$ is PDS. Similarly, $\lim_{N \rightarrow \infty} E(Z_i' Z_i)$ is PDS since $\lim_{N \rightarrow \infty} E(Z_i^{AB'} Z_i^{AB}) = \text{diag}(\sigma_y^2 I_{T-2}, \sigma^2 I_{T-3}, \sigma^2 I_{T-4}, \dots, \sigma^2 I_1)$ is PDS.

Parts d) and e): When $\rho = 1$, $v_i = \varepsilon_i$, $E[(\Delta v_i' \ v_i')' (\Delta v_i' \ v_i')] = \sigma^2 A$, $\Delta y_i^t = \varepsilon_i^t$ and $Z_i^S = \text{diag}(y_{i,1} I_{T-2}, \text{diag}(\varepsilon_i^2, \dots, \varepsilon_i^{T-1}))$. It follows that $\lim_{N \rightarrow \infty} E(Z_i^{S'} A Z_i^S) = \text{diag}(\sigma_y^2 H, \sigma^2 I_m)$, which is PDS, $\text{plim}_{N \rightarrow \infty} W_{N,Arbov1} = \sigma^2 \times \text{plim}_{N \rightarrow \infty} W_{N,Arbov2}(\hat{\rho}_1) = \sigma^{-2} I_m$ and $\text{plim}_{N \rightarrow \infty} W_{N,SY S1} = \sigma^2 \times \text{plim}_{N \rightarrow \infty} W_{N,SY S2}(\hat{\rho}_1) = \text{diag}(\sigma_y^{-2} H^{-1}, \sigma^{-2} I_m)$.

Part f): These results follow from parts a) and b), that is, from $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N Z_i^{I'} \Delta y_{i,-1} = \text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N Z_i^{I'} (\varepsilon_i^{T-1})' = 0$, while $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N Z_i^{II'} y_{i,-1} = \sigma^2 \iota_m = \check{X}_8 \neq 0$, and from the fact that $N^{-1/2} \sum_{i=1}^N Z_i^{II'} v_i \xrightarrow{d} \check{X}_7$ and $N^{-1/2} \sum_{i=1}^N Z_i^{I'} \Delta v_i = N^{-1/2} \sum_{i=1}^N K^{AB} \times [Z_i^I \ Z_i^D]' \Delta v_i \xrightarrow{d} \check{X}_5$, where K^{AB} is a constant and nonsingular matrix.

Part g): From parts b), d) and e) we have $\check{X}_8 = \sigma^2 \iota_m$, $Var(\check{X}_7) = W_{SYS2,22}^{-1} = W_{Arbov2}^{-1} = \sigma^4 I_m$ and $W_{SYS2,12} = W_{SYS2,21} = 0$. It follows that $Asyvar(\hat{\rho}_{SYS2}) = (\check{X}'_8 W_{SYS2,22} \check{X}_8)^{-1} = (\check{X}'_8 W_{Arbov2} \check{X}_8)^{-1} = Asyvar(\hat{\rho}_{Arbov2}) = 1/m$.

Part h): Recall that $\hat{\rho}_{AB1} = \{\sum_{t=2}^{T-1} [\tilde{y}'_{-1,t-1} Z_t (Z'_t Z_t)^{-1} Z'_t \tilde{y}_{-1,t-1}]\}^{-1} \sum_{t=2}^{T-1} [\tilde{y}'_{-1,t-1} Z_t (Z'_t Z_t)^{-1} \times Z'_t \tilde{y}_t]$. Since $\sum_{t=2}^{T-1} [\tilde{y}'_{-1,t-1} Z_t (Z'_t Z_t)^{-1} Z'_t \tilde{v}_t] = (\sum_{i=1}^N \Delta y'_{i,-1} Z_i) [\sum_{i=1}^N (Z'_i H Z_i)]^{-1} \sum_{i=1}^N Z'_i \Delta v_i$ and $\sum_{t=2}^{T-1} [\tilde{y}'_{-1,t-1} Z_t (Z'_t Z_t)^{-1} Z'_t \tilde{y}_{-1,t-1}] = (\sum_{i=1}^N \Delta y'_{i,-1} Z_i) [\sum_{i=1}^N (Z'_i H Z_i)]^{-1} \sum_{i=1}^N Z'_i \Delta y_{i,-1}$, it follows from parts a) and c) that the numerator and denominator of $\hat{\rho}_{AB1} - \rho$ converge in distribution.

We now show that $\lim_{N \rightarrow \infty} E[\tilde{y}'_{-1,t-1} Z_t (Z'_t Z_t)^{-1} Z'_t \tilde{v}_t] < 0$ for $t = 2, \dots, T-1$. Note that $\lim_{N \rightarrow \infty} E[\tilde{y}'_{-1,t-1} Z_t (Z'_t Z_t)^{-1} Z'_t \tilde{v}_t] = \lim_{N \rightarrow \infty} E\{\sum_{i=1}^N [\tilde{y}_{i,-1,t-1} z'_{i,t} (Z'_t Z_t)^{-1} z_{i,t} \tilde{\varepsilon}_{i,t}]\} = \lim_{N \rightarrow \infty} \sum_{i=1}^N [E(z'_{i,t} (Z'_t Z_t)^{-1} z_{i,t}) \times E(\tilde{y}_{i,-1,t-1} \tilde{\varepsilon}_{i,t})]$.

Now, $E(\tilde{y}_{i,-1,t-1} \tilde{\varepsilon}_{i,t}) = \left(\frac{T-t}{T-t+1}\right) E\left[(y_{i,t-1} - \frac{1}{T-t} \sum_{s=t}^{T-1} y_{i,s})(\varepsilon_{i,t} - \frac{1}{T-t} \sum_{s=t+1}^T \varepsilon_{i,s})\right] = \left(\frac{1}{T-t+1}\right) E\left[-\sum_{k=0}^{T-1-t} \rho^k \varepsilon_{i,t}^2 + \frac{1}{T-t} \sum_{s=t+1}^{T-1} \sum_{k=0}^{T-1-s} \rho^k \varepsilon_{i,s}^2\right] = -\sigma^2 \left(\frac{1}{T-t+1}\right) \left(\frac{1}{T-t}\right) \times \left[\sum_{s=t+1}^T \sum_{k=T-1-s+1}^{T-1-t} \rho^k\right]$. It follows that $\lim_{N \rightarrow \infty} E(\tilde{y}_{i,-1,t-1} \tilde{\varepsilon}_{i,t}) = \lim_{\rho \uparrow 1} E(\tilde{y}_{i,-1,t-1} \tilde{\varepsilon}_{i,t}) = -\frac{1}{2} \sigma^2$. We also have $\lim_{N \rightarrow \infty} E(\sum_{i=1}^N z'_{i,t} (Z'_t Z_t)^{-1} z_{i,t}) = E[t\rho(I_{t-1})] = t-1 > 0$.

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

To complete the proof of the asymptotic biasedness of $\hat{\rho}_{AB}$, we write the numerator of $\hat{\rho}_{AB} - \rho$, i.e. $\sum_{t=2}^{T-1} [(\sum_{i=1}^N \tilde{y}_{i,-1,t-1} z'_{i,t}) (Z'_t Z_t)^{-1} (\sum_{i=1}^N z_{i,t} \tilde{\varepsilon}_{i,t})]$, as the sum of two terms:

$$\left\{ \sum_{t=2}^{T-1} \left[\sum_{i=1}^N z'_{i,t} (Z'_t Z_t)^{-1} z_{i,t} \left(\frac{1}{T-t+1}\right) \left(-\sum_{k=0}^{T-1-t} \rho^k \varepsilon_{i,t}^2 + \frac{1}{T-t} \sum_{s=t+1}^{T-1} \sum_{k=0}^{T-1-s} \rho^k \varepsilon_{i,s}^2\right) \right] \right\} \quad (15)$$

$$+ \left\{ \sum_{t=2}^{T-1} [\tilde{y}'_{-1,t-1} Z_t (Z'_t Z_t)^{-1} Z'_t \tilde{\varepsilon}_t - \sum_{i=1}^N z'_{i,t} (Z'_t Z_t)^{-1} z_{i,t} \left(\frac{1}{T-t+1}\right) \left(-\sum_{k=0}^{T-1-t} \rho^k \varepsilon_{i,t}^2 + \frac{1}{T-t} \sum_{s=t+1}^{T-1} \sum_{k=0}^{T-1-s} \rho^k \varepsilon_{i,s}^2\right)] \right\}. \quad (16)$$

The first term, (15), converges in probability to $\lim_{N \rightarrow \infty} \sum_{t=2}^{T-1} E[\tilde{y}'_{-1,t-1} Z_t (Z'_t Z_t)^{-1} Z'_t \tilde{\varepsilon}_t]$. Since the numerator of $\hat{\rho}_{AB} - \rho$, $\sum_{t=2}^{T-1} \tilde{y}'_{-1,t-1} Z_t (Z'_t Z_t)^{-1} Z'_t \tilde{\varepsilon}_t$, converges in distribution to a random variable with mean $\lim_{N \rightarrow \infty} \sum_{t=2}^{T-1} E[\tilde{y}'_{-1,t-1} Z_t (Z'_t Z_t)^{-1} Z'_t \tilde{\varepsilon}_t]$, the second term, (16), converges in distribution to a random variable with mean zero. For convenience we will assume that the $\varepsilon'_{i,t} s$ are symmetrically distributed around zero. Doing so does not entail a

loss of generality since imposing this assumption does not affect the asymptotic distribution of $\widehat{\rho}_{AB}$. Noting that the second term does not involve higher powers of $\varepsilon_{i,t}$ (higher than one) and using that the $\varepsilon'_{i,t}$ s are symmetrically distributed around zero, it follows that the second term is symmetrically distributed around zero and also asymptotically uncorrelated with the denominator of $\widehat{\rho}_{AB} - \rho$, $\sum_{t=2}^{T-1} [\widetilde{y}'_{-1,t-1} Z_t (Z'_t Z_t)^{-1} Z'_t \widetilde{y}_{-1,t-1}]$. We conclude that the expectation of the second term divided by the denominator converges to zero as $N \rightarrow \infty$. The ratio of the first term and the denominator converges to a negative constant divided by a positive random variable and gives rise to the negative bias of $\widehat{\rho}_{AB}$. \square

Lemma 8 *Let $\{y_{i,t}\}$ be a stationary process and let $u_{i,t} = y_{i,t} - \mu_i$.*

$$\begin{aligned} \text{Then } E(u_{i,t-1} \Delta u_{i,t}) &= -\frac{\sigma^2}{1+\rho}, \quad E[(\Delta u_{i,t})^2] = \frac{2\sigma^2}{1+\rho}, \quad E(u_{i,t}^4) = \frac{(1-\rho^2)\kappa+6\rho^2}{(1-\rho^2)^2(1+\rho^2)}\sigma^4, \\ E[(\Delta u_{i,t})^4] &= \frac{(1-\rho)^2[\kappa(1-\rho^2)+6\rho^2]}{(1+\rho)^2(1+\rho^2)}\sigma^4 + \frac{6(1-\rho)}{1+\rho}\sigma^4 + \kappa\sigma^4, \quad E[u_{i,t-1}(\Delta u_{i,t})^3] = \\ &= -\frac{(1-\rho)[\kappa(1-\rho^2)+6\rho^2]}{(1+\rho)^2(1+\rho^2)}\sigma^4 - \frac{3}{(1+\rho)}\sigma^4, \quad \text{and } E[u_{i,t-1}^2(\Delta u_{i,t})^2] = \frac{(\kappa(1-\rho^2)+6\rho^2)}{(1+\rho)^2(1+\rho^2)}\sigma^4 + \frac{1}{1-\rho^2}\sigma^4. \end{aligned}$$

Proof of lemma 8: Note that $u_{i,t} = \rho u_{i,t-1} + \varepsilon_{i,t}$ and $\Delta u_{i,t} = (\rho - 1)u_{i,t-1} + \varepsilon_{i,t}$. Moreover $\{u_{i,t}\}$ is a stationary process. Then verification of the first two claims is straightforward. The other claims are proved as follows:

$$\begin{aligned} E(u_{i,t}^4) &= E(\rho^4 u_{i,t-1}^4 + 6\rho^2 u_{i,t-1}^2 \varepsilon_{i,t}^2 + \varepsilon_{i,t}^4) = \rho^4 E(u_{i,t-1}^4) + \frac{6\sigma^4 \rho^2}{(1-\rho^2)} + \kappa\sigma^4 \Leftrightarrow \\ E(u_{i,t}^4) &= \frac{\kappa\sigma^4}{1-\rho^4} + \frac{6\sigma^4 \rho^2}{(1-\rho^2)(1-\rho^4)} = \frac{(1-\rho^2)\kappa+6\rho^2}{(1-\rho^2)^2(1+\rho^2)}\sigma^4, \\ E[(\Delta u_{i,t})^4] &= E[(\rho-1)^4 u_{i,t-1}^4 + 6(\rho-1)^2 u_{i,t-1}^2 \varepsilon_{i,t}^2 + \varepsilon_{i,t}^4] = \frac{(1-\rho)^2[\kappa(1-\rho^2)+6\rho^2]}{(1+\rho)^2(1+\rho^2)}\sigma^4 + \frac{6(1-\rho)}{1+\rho}\sigma^4 + \kappa\sigma^4, \\ E[u_{i,t-1}(\Delta u_{i,t})^3] &= E[(\rho-1)^3 u_{i,t-1}^4 + 3(\rho-1)u_{i,t-1}^2 \varepsilon_{i,t}^2] = -\frac{(1-\rho)[\kappa(1-\rho^2)+6\rho^2]}{(1+\rho)^2(1+\rho^2)}\sigma^4 - \frac{3}{(1+\rho)}\sigma^4, \quad \text{and} \\ E[u_{i,t-1}^2(\Delta u_{i,t})^2] &= E[(\rho-1)^2 u_{i,t-1}^4 + u_{i,t-1}^2 \varepsilon_{i,t}^2] = \frac{(\kappa(1-\rho^2)+6\rho^2)}{(1+\rho)^2(1+\rho^2)}\sigma^4 + \frac{1}{1-\rho^2}\sigma^4. \quad \square \end{aligned}$$

Lemma 9 *Let $\{y_{i,t}\}$ be a stationary process. Then*

$$\begin{aligned} E(y_{i,1} \Delta y_{i,2}) &= -\frac{\sigma^2}{1+\rho}, \quad E(y_{i,2} \Delta y_{i,2}) = \frac{\sigma^2}{1+\rho}, \\ \text{Var}(y_{i,1} \Delta \varepsilon_{i,3}) &= \frac{2\sigma^2[\sigma^2 + \sigma_\mu^2(1-\rho^2)]}{(1+\rho)(1-\rho)}, \quad \text{Var}[(\varepsilon_{i,3} + (1-\rho)\mu_i) \Delta y_{i,2}] = \frac{2\sigma^2[\sigma^2 + \sigma_\mu^2(1-\rho^2)]}{1+\rho}, \\ \text{Var}(y_{i,1} \Delta y_{i,2}) &= \frac{(\kappa(1-\rho^2)+6\rho^2)\sigma^4}{(1+\rho)^2(1+\rho^2)} + \frac{\sigma^4}{(1+\rho)(1-\rho)} + \frac{\sigma^2[2\sigma_\mu^2(1+\rho) - \sigma^2]}{(1+\rho)^2}, \\ \text{Var}(y_{i,2} \Delta y_{i,2}) &= \frac{\rho^2(\kappa(1-\rho^2)+6\rho^2)}{(1+\rho)^2(1+\rho^2)}\sigma^4 + \frac{\sigma^4}{(1+\rho)(1-\rho)} + \frac{\kappa(1+\rho)^2 - 1 - 6\rho(1+\rho) + 2(\sigma_\mu^2/\sigma^2)(1+\rho)}{(1+\rho)^2}\sigma^4, \\ \text{Cov}(y_{i,1} \Delta \varepsilon_{i,3}, y_{i,1} \Delta y_{i,2}) &= -\frac{\sigma^2[\sigma^2 + \sigma_\mu^2(1-\rho^2)]}{(1-\rho)(1+\rho)}, \quad \text{and} \\ \text{Cov}[(\varepsilon_{i,3} + (1-\rho)\mu_i) \Delta y_{i,2}, y_{i,2} \Delta y_{i,2}] &= \frac{2\sigma^2 \sigma_\mu^2 (1-\rho)}{1+\rho}. \end{aligned}$$

Proof of lemma 9: Noting that $y_{i,t} = u_{i,t} + \mu_i$, application of lemma 8 yields

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

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

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

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

$$\begin{aligned} \text{Var}(y_{i,1}\Delta y_{i,2}) &= E[(u_{i,1}\Delta u_{i,2} + \mu_i\Delta u_{i,2} + \frac{\sigma^2}{1+\rho})^2] = E[(u_{i,1}\Delta u_{i,2})^2 - \frac{2\sigma^4}{(1+\rho)^2} + \frac{2\sigma^2\sigma_\mu^2}{1+\rho} + \frac{\sigma^4}{(1+\rho)^2}] = \\ &= \frac{(\kappa(1-\rho^2)+6\rho^2)\sigma^4}{(1+\rho)^2(1+\rho^2)} + \frac{\sigma^4}{(1+\rho)(1-\rho)} + \frac{\sigma^2[2\sigma_\mu^2(1+\rho)-\sigma^2]}{(1+\rho)^2}, \end{aligned}$$

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

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

$$\begin{aligned} &= \frac{(\kappa(1-\rho^2)+6\rho^2)\sigma^4}{(1+\rho)^2(1+\rho^2)}\sigma^4 + \frac{\sigma^4}{1-\rho^2} + \frac{(1-\rho)^2[\kappa(1-\rho^2)+6\rho^2]\sigma^4}{(1+\rho)^2(1+\rho^2)} + \frac{6(1-\rho)\sigma^4}{1+\rho} + \kappa\sigma^4 + \frac{2\sigma^2\sigma_\mu^2}{1+\rho} + \frac{\sigma^4}{(1+\rho)^2} - \frac{2(1-\rho)[\kappa(1-\rho^2)+6\rho^2]\sigma^4}{(1+\rho)^2(1+\rho^2)} \\ &- \frac{6}{(1+\rho)}\sigma^4 + \frac{2\sigma^4}{(1+\rho)^2} - \frac{4\sigma^4}{(1+\rho)^2} = \frac{\rho^2(\kappa(1-\rho^2)+6\rho^2)\sigma^4}{(1+\rho)^2(1+\rho^2)} + \frac{\sigma^4}{(1+\rho)(1-\rho)} + \frac{\kappa(1+\rho)^2 - 1 - 6\rho(1+\rho) + 2(\sigma_\mu^2/\sigma^2)(1+\rho)}{(1+\rho)^2}\sigma^4, \end{aligned}$$

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

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

$$\text{Cov}[(\varepsilon_{i,3} + (1-\rho)\mu_i)\Delta y_{i,2}, y_{i,2}\Delta y_{i,2}] = E[(1-\rho)(\mu_i\Delta u_{i,2})^2] = \frac{2\sigma^2\sigma_\mu^2(1-\rho)}{1+\rho}. \quad \square$$

A.2 Proof of theorem 2

When $T = 3$, $\hat{\rho}_{AB} = \rho + \frac{N^{-1}\sum_{i=1}^N y_{i,1}\Delta \varepsilon_{i,3}}{N^{-1}\sum_{i=1}^N y_{i,1}\Delta y_{i,2}}$ and $\hat{\rho}_{Arbov} = \rho + \frac{N^{-1}\sum_{i=1}^N (\varepsilon_{i,3} + (1-\rho)\mu_i)\Delta y_{i,2}}{N^{-1}\sum_{i=1}^N y_{i,2}\Delta y_{i,2}}$.

Let us define $X_1 = N^{-1}\sum_{i=1}^N y_{i,1}\Delta \varepsilon_{i,3}$, $X_2 = N^{-1}\sum_{i=1}^N y_{i,1}\Delta y_{i,2}$, $X_3 = N^{-1/2}\sum_{i=1}^N (\varepsilon_{i,3} + (1-\rho)\mu_i)\Delta y_{i,2}$ and $X_4 = N^{-1}\sum_{i=1}^N y_{i,2}\Delta y_{i,2}$.

Let the CS model hold. Then using the results in lemma 9 we obtain as $N \rightarrow \infty$ for the parameter sequence $\rho = 1 - \lambda N^{-1}$ that $X_1 \xrightarrow{d} N(0, \sigma^4/\lambda)$, $X_2 \xrightarrow{d} N(-\sigma^2/2, \sigma^4/(2\lambda))$, $X_3 \xrightarrow{d} N(0, \sigma^4)$, and $X_4 \xrightarrow{d} N(\sigma^2/2, \sigma^4/(2\lambda))$. It is also easily verified that $\lim_{N \rightarrow \infty} \text{Cov}(X_1, X_2) = -\sigma^4/(2\lambda)$ and $\lim_{N \rightarrow \infty} \text{Cov}(X_3, X_4) = 0$. It follows that

$$\hat{\rho}_{AB} \xrightarrow{d} \rho + \frac{\tilde{X}_1}{\tilde{X}_2}, \text{ where } \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix} \sim N \left[-\frac{1}{2}\sqrt{\lambda} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \right], \text{ and}$$

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

Lemma 10 Let $\{y_{i,t}\}$ be a stationary process, let $u_{i,t} = y_{i,t} - \mu_i$, let l, p , and q be positive integers, and let $\kappa_q = |E(\varepsilon_{i,t}^q)/\sigma^q| < \infty$.

Then (a) $\lim_{\rho \uparrow 1} (1-\rho)^p E(u_{i,t}^{2p}) < \infty$, (b) $\lim_{\rho \uparrow 1} (1-\rho)^p |E(u_{i,t}^{2p+1})| < \infty$, (c) $\lim_{\rho \uparrow 1} E[(\Delta u_{i,t})^p] = \kappa_p \sigma^p$, (d) $\lim_{\rho \uparrow 1} (1-\rho)^p |E[u_{i,s}^{2p} (\Delta u_{i,t})^q]| < \infty$, (e) $\lim_{\rho \uparrow 1} (1-\rho)^p |E[u_{i,s}^{2p+1} (\Delta u_{i,t})^q]| < \infty$, (f) $\lim_{\rho \uparrow 1} (1-\rho)^p |E[\varepsilon_{i,k}^l u_{i,s}^{2p} (\Delta u_{i,t})^q]| < \infty$, and (g) $\lim_{\rho \uparrow 1} (1-\rho)^p |E[\varepsilon_{i,k}^l u_{i,s}^{2p+1} (\Delta u_{i,t})^q]| < \infty$.

Proof of lemma 10: Note that $\{u_{i,t}\}$ is a stationary process, $u_{i,t} = \rho u_{i,t-1} + \varepsilon_{i,t}$ and $\Delta u_{i,t} = (\rho - 1)u_{i,t-1} + \varepsilon_{i,t}$. Moreover, $E(u_{i,t}^p) = E[(\rho u_{i,t-1} + \varepsilon_{i,t})^p] = E[\sum_{k=0}^p B(p, k)(\rho u_{i,t-1})^{p-k} \varepsilon_{i,t}^k]$, where $B(p, k) = [k!(p-k)!]^{-1} p!$.

We prove (a) and (b) together. The proof proceeds by induction:

First consider $p = 0$: $E(u_{i,t}^0) = 1$ and $E(u_{i,t}^1) = 0$ and hence $\lim_{\rho \uparrow 1} E(u_{i,t}^0) < \infty$ and $\lim_{\rho \uparrow 1} |E(u_{i,t}^1)| < \infty$.

Now let $p > 0$ and suppose that $\lim_{\rho \uparrow 1} (1-\rho)^q E(u_{i,t}^{2q}) < \infty$ and $\lim_{\rho \uparrow 1} (1-\rho)^q |E(u_{i,t}^{2q+1})| < \infty$ for $q \leq p-1$. Note that $\lim_{\rho \uparrow 1} (1-\rho^{2p})^{-1}(1-\rho) = \lim_{\rho \uparrow 1} (\sum_{k=0}^{2p-1} \rho^k)^{-1} = 1/(2p)$ and $\lim_{\rho \uparrow 1} (1-\rho^{2p+1})^{-1}(1-\rho) = 1/(2p+1)$. It follows that $\lim_{\rho \uparrow 1} (1-\rho)^p E(u_{i,t}^{2p}) = \lim_{\rho \uparrow 1} (1-\rho^{2p})^{-1}(1-\rho)^p B(2p, 2) E(u_{i,t}^{2p-2}) E(\varepsilon_{i,t}^2) = \sigma^2 (2p)^{-1} B(2p, 2) \times \lim_{\rho \uparrow 1} (1-\rho)^{p-1} E(u_{i,t}^{2p-2}) < \infty$. It also follows that $\lim_{\rho \uparrow 1} (1-\rho)^p |E(u_{i,t}^{2p+1})| \leq \lim_{\rho \uparrow 1} (1-\rho^{2p+1})^{-1}(1-\rho)^p B(2p+1, 2) \times |E(u_{i,t}^{2p-1})| E(\varepsilon_{i,t}^2) + \lim_{\rho \uparrow 1} (1-\rho^{2p+1})^{-1}(1-\rho)^p B(2p+1, 3) E(u_{i,t}^{2p-2}) |E(\varepsilon_{i,t}^3)| = \sigma^2 (2p+1)^{-1} B(2p+1, 2) \times \lim_{\rho \uparrow 1} (1-\rho)^{p-1} |E(u_{i,t}^{2p-1})| + |\kappa_3| \sigma^3 (2p+1)^{-1} B(2p+1, 3) \times \lim_{\rho \uparrow 1} (1-\rho)^{p-1} E(u_{i,t}^{2p-2}) < \infty$.

The proofs of (c) to (g) are now straightforward:

$$(c): \lim_{\rho \uparrow 1} E[(\Delta u_{i,t})^p] = \lim_{\rho \uparrow 1} E[(\rho - 1)u_{i,t-1} + \varepsilon_{i,t}]^p = E[(\varepsilon_{i,t})^p] = \kappa_p \sigma^p,$$

$$(d): \lim_{\rho \uparrow 1} (1-\rho)^p |E[u_{i,s}^{2p} (\Delta u_{i,t})^q]| = \lim_{\rho \uparrow 1} (1-\rho)^p |E[u_{i,s}^{2p} \varepsilon_{i,t}^q]| < \infty,$$

$$(e): \lim_{\rho \uparrow 1} (1-\rho)^p |E[u_{i,s}^{2p+1} (\Delta u_{i,t})^q]| = \lim_{\rho \uparrow 1} (1-\rho)^p |E[u_{i,s}^{2p+1} (B(q, 1)(\rho - 1)u_{i,t-1} \varepsilon_{i,t}^{q-1} + \varepsilon_{i,t}^q)]| < \infty,$$

$$(f): \lim_{\rho \uparrow 1} (1-\rho)^p |E[\varepsilon_{i,k}^l u_{i,s}^{2p} (\Delta u_{i,t})^q]| = \lim_{\rho \uparrow 1} (1-\rho)^p |E[\varepsilon_{i,k}^l u_{i,s}^{2p} \varepsilon_{i,t}^q]| < \infty, \text{ and finally}$$

$$(g): \lim_{\rho \uparrow 1} (1-\rho)^p |E[\varepsilon_{i,k}^l u_{i,s}^{2p+1} (\Delta u_{i,t})^q]| = \lim_{\rho \uparrow 1} (1-\rho)^p |E[\varepsilon_{i,k}^l u_{i,s}^{2p+1} (B(q, 1)(\rho - 1)u_{i,t-1} \varepsilon_{i,t}^{q-1} + \varepsilon_{i,t}^q)]| < \infty. \quad \square$$

A.3 Proof of theorem 3

Part a): The proof of the first two results is similar to the proof of the first part of theorem 2:

For $\rho = 1 - \lambda N^{-1} \lim_{N \rightarrow \infty} E(y_{i,1} \Delta y_{i,t-1}) = \lim_{N \rightarrow \infty} \rho^{t-3} E(y_{i,1} \Delta y_{i,2}) = -\frac{1}{2} \sigma^2$ and $\lim_{N \rightarrow \infty} N^{-1} Var(y_{i,1} \Delta v_{i,t}) = \lim_{N \rightarrow \infty} N^{-1} E[(y_{i,1} \Delta \varepsilon_{i,t})^2] = \lim_{N \rightarrow \infty} N^{-1} E[(y_{i,1} \Delta \varepsilon_{i,3})^2] = \lim_{N \rightarrow \infty} 2\sigma^2 N^{-1} E(y_{i,1}^2) = \lim_{N \rightarrow \infty} 2\sigma^2 N^{-1} E(u_{i,1}^2) = \sigma^4/\lambda$.

Note that $Var(y_{i,1} \Delta y_{i,t-1}) = E[(u_{i,1} \Delta u_{i,t-1} + \mu_i \Delta u_{i,t-1} + \frac{\sigma^2}{1+\rho} \rho^{t-3})^2]$, $u_{i,t} = \rho u_{i,t-1} + \varepsilon_{i,t}$ and $\Delta u_{i,t} = (\rho - 1)u_{i,t-1} + \varepsilon_{i,t}$. Then it is easily seen that for $\rho = 1 - \lambda N^{-1}$ $\lim_{N \rightarrow \infty} N^{-1} Var(y_{i,1} \Delta y_{i,t-1}) = \lim_{N \rightarrow \infty} N^{-1} E[(u_{i,1} \Delta u_{i,t-1})^2] = \lim_{N \rightarrow \infty} N^{-1} E\{[(\rho-1)u_{i,1}^2 + u_{i,1}\varepsilon_{i,t-1}]^2\} = \lim_{N \rightarrow \infty} N^{-1} E[(u_{i,1}\varepsilon_{i,t-1})^2] = \sigma^4/2\lambda$; that if $s < t - 1$ and $\rho = 1 - \lambda N^{-1}$ $\lim_{N \rightarrow \infty} N^{-1} Cov(y_{i,1} \Delta y_{i,s}, y_{i,1} \Delta y_{i,t-1}) = \lim_{N \rightarrow \infty} N^{-1} E(u_{i,1}^2 \Delta u_{i,s} \Delta u_{i,t-1}) = \lim_{N \rightarrow \infty} N^{-1} E(u_{i,1}^2 \Delta u_{i,s} \Delta u_{i,s+1}) = \lim_{N \rightarrow \infty} N^{-1} E\{u_{i,1}^2 [(\rho-1)u_{s-1} + \varepsilon_s][(\rho-1)(\rho u_{s-1} + \varepsilon_s)]\} = 0$; $\lim_{N \rightarrow \infty} N^{-1} Cov(y_{i,1} \Delta v_{i,t}, y_{i,1} \Delta y_{i,t-1}) = \lim_{N \rightarrow \infty} N^{-1} E(y_{i,1}^2 \Delta \varepsilon_{i,t} \Delta u_{i,t-1}) = -\sigma^4/2\lambda$ and finally that $\lim_{N \rightarrow \infty} N^{-1} Cov(y_{i,1} \Delta v_{i,t}, y_{i,1} \Delta y_{i,t}) = \lim_{N \rightarrow \infty} N^{-1} E(y_{i,1}^2 \Delta \varepsilon_{i,t} \Delta u_{i,t}) = \sigma^4/2\lambda$.

Thus for $\rho = 1 - \lambda N^{-1} \lim_{N \rightarrow \infty} N^{-1} E(Z_i^{I'} \Delta v_i \Delta v_i' Z_i^I) = \lim_{N \rightarrow \infty} \sigma^2 N^{-1} E(y_{i,1}^2) H = (\sigma^4/2\lambda)H$, $\lim_{N \rightarrow \infty} N^{-1} E(Z_i^{I'} \Delta y_{i,-1} \Delta y_{i,-1}' Z_i^I) = \lim_{N \rightarrow \infty} \sigma^2 N^{-1} E(y_{i,1}^2) I = (\sigma^4/2\lambda)I$ and $\lim_{N \rightarrow \infty} N^{-1} E(Z_i^{I'} \Delta y_{i,-1} \Delta v_i' Z_i^I) = -\lim_{N \rightarrow \infty} \sigma^2 N^{-1} E(y_{i,1}^2) C = -(\sigma^4/2\lambda)C$ and hence $\begin{bmatrix} N^{-1} \sum_{i=1}^N Z_i^{I'} \Delta v_i \\ N^{-1} \sum_{i=1}^N Z_i^{I'} \Delta y_{i,-1} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \tilde{X}_{51} \\ \tilde{X}_{61} \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ -\frac{1}{2} \sigma^2 \iota \end{pmatrix}, \left(\frac{\sigma^4}{2\lambda}\right) \begin{pmatrix} H & -C' \\ -C & I \end{pmatrix} \right]$.

Note that $\tilde{\Sigma}_{56,11} = E(\tilde{X}_{51} \tilde{X}_{61}') = -(\sigma^4/2\lambda)C' \neq 0$. Moreover, when $T = 3$, $\tilde{X}_{51} = \tilde{X}_1/\sqrt{\lambda}$ and $\tilde{X}_{61} = \tilde{X}_2/\sqrt{\lambda}$, where \tilde{X}_1 and \tilde{X}_2 are defined in theorem 2.

Consider now the scaled sums $N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta v_i$ and $N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta y_{i,-1}$. For $\rho = 1 - \lambda N^{-1}$ we obtain $\lim_{N \rightarrow \infty} N^{1/2} E(Z_i^{D'} \Delta y_{i,-1}) = \lim_{N \rightarrow \infty} N^{1/2} E(Z_i^{D'} \Delta u_{i,-1}) = \lim_{N \rightarrow \infty} N^{1/2} E\{Z_i^{D'} [(\rho - 1)u_{i,-2} + \varepsilon_{i,-1}]\} = 0$. In addition, $\lim_{N \rightarrow \infty} E(Z_i^{D'} \Delta v_i \Delta v_i' Z_i^D) = \lim_{N \rightarrow \infty} \sigma^2 E(Z_i^{D'} H Z_i^D)$, $\lim_{N \rightarrow \infty} E(Z_i^{D'} \Delta y_{i,-1} \Delta y_{i,-1}' Z_i^D) = \lim_{N \rightarrow \infty} \sigma^2 E(Z_i^{D'} Z_i^D)$ and $\lim_{N \rightarrow \infty} E(Z_i^{D'} \Delta y_{i,-1} \Delta v_i' Z_i^D) = -\lim_{N \rightarrow \infty} \sigma^2 E(Z_i^{D'} C Z_i^D)$. Hence for $\rho = 1 - \lambda N^{-1}$

$$\begin{bmatrix} N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta v_i \\ N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta y_{i,-1} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \tilde{X}_{52} \\ \tilde{X}_{62} \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^4 \begin{pmatrix} \text{diag}(H_{T-3}, H_{T-4}, \dots, H_1) & -\text{diag}(C_{T-3}, C_{T-4}, \dots, C_1)' \\ -\text{diag}(C_{T-3}, C_{T-4}, \dots, C_1) & I \end{pmatrix} \right].$$

Note that $\tilde{\Sigma}_{56,22} = E(\tilde{X}_{52} \tilde{X}_{62}') = -\sigma^4 \text{diag}(C_{T-3}, C_{T-4}, \dots, C_1)' \neq 0$.

Next consider the off-diagonal blocks $\tilde{\Sigma}_{55,12}$, $\tilde{\Sigma}_{66,12}$, $\tilde{\Sigma}_{56,12}$ and $\tilde{\Sigma}_{56,21}$. Since $\lim_{N \rightarrow \infty} E(Z_i^{I'} \Delta v_i \Delta v_i' Z_i^D) = \lim_{N \rightarrow \infty} \sigma^2 E(Z_i^{I'} H Z_i^D) = \lim_{N \rightarrow \infty} \sigma^2 H E(y_{i,1} Z_i^D) = \lim_{N \rightarrow \infty} \sigma^2 H E(y_{i,1} [d_1 \Delta y_{i,2} \ d_2 \rho \Delta y_{i,2} \ \dots \ d_{T-3} \rho^{T-4} \Delta y_{i,2}]) = \lim_{N \rightarrow \infty} -\frac{1}{2} \sigma^4 H [d_1 \ d_2 \ \dots \ d_{T-3}]$, we have $\tilde{\Sigma}_{55,12} = E(\tilde{X}_{51} \tilde{X}'_{52}) = \lim_{N \rightarrow \infty} \sigma^2 N^{-1/2} E(Z_i^{I'} H Z_i^D) = 0$. Similarly, $\tilde{\Sigma}_{66,12} = E(\tilde{X}_{61} \tilde{X}'_{62}) = \lim_{N \rightarrow \infty} \sigma^2 N^{-1/2} E(Z_i^{I'} Z_i^D) = 0$, $\tilde{\Sigma}_{56,12} = E(\tilde{X}_{51} \tilde{X}'_{62}) = -\lim_{N \rightarrow \infty} \sigma^2 \times N^{-1/2} E(Z_i^{I'} C' Z_i^D) = 0$, and $\tilde{\Sigma}_{56,21} = E(\tilde{X}_{52} \tilde{X}'_{61}) = -\lim_{N \rightarrow \infty} \sigma^2 N^{-1/2} E(Z_i^{D'} C' Z_i^I) = 0$.

Recall that $Z_i' = K^{AB} Z_i^{AB'}$, where $Z_i^{AB} = [Z_i^I \ Z_i^D]$ and $\text{rank}(K^{AB}) = m$. Therefore $\hat{\rho}_{AB1}$ is equal to an GMM estimator that exploits $E(Z_i^{AB'} \Delta v_i) = 0$ and uses the weight matrix $(N^{-1} \sum_{i=1}^N Z_i^{AB'} H Z_i^{AB})^{-1}$. Consider the local-to-unity limiting behaviour of this weight matrix. Since $\lim_{N \rightarrow \infty} N^{-2} \text{Var}(y_{i,1}^2) < \infty$, $\lim_{N \rightarrow \infty} N^{-1} \text{Var}(y_{i,1} \Delta y_{i,t}) < \infty$ and $\lim_{N \rightarrow \infty} \text{Var}([\Delta y_{i,t}]^2) < \infty$, it is easily seen that for $\rho = 1 - \lambda N^{-1}$ $\sigma^2 N^{-2} \sum_{i=1}^N (Z_i^{I'} H Z_i^I) \xrightarrow{q.m.} \tilde{\Sigma}_{55,11}$ and $\sigma^2 N^{-3/2} \sum_{i=1}^N (Z_i^{I'} H Z_i^D) \xrightarrow{q.m.} 0$. Moreover $\sigma^2 N^{-1} \sum_{i=1}^N (Z_i^{D'} H Z_i^D) \xrightarrow{q.m.} \tilde{\Sigma}_{55,22}$. Therefore $\text{plim}_{N \rightarrow \infty} \sigma^2 N^{-2} \sum_{i=1}^N (Z_i^{I'} H Z_i^I) = \tilde{\Sigma}_{55,11}$, $\text{plim}_{N \rightarrow \infty} \sigma^2 N^{-1} \sum_{i=1}^N (Z_i^{D'} H Z_i^D) = \tilde{\Sigma}_{55,22}$ and $\text{plim}_{N \rightarrow \infty} \sigma^2 N^{-3/2} \sum_{i=1}^N (Z_i^{I'} H Z_i^D) = \tilde{\Sigma}_{55,12} = 0$. Furthermore $\tilde{\Sigma}_{55}$ is PDS. It follows from the above results that $\hat{\rho}_{AB1} - \rho \xrightarrow{d} \tilde{X}'_6 \tilde{\Sigma}_{55}^{-1} \tilde{X}_5 / \tilde{X}'_6 \tilde{\Sigma}_{55}^{-1} \tilde{X}_6$. Assuming that $\hat{\rho}_{AB}$ exploits $E(Z_i^{AB'} \Delta v_i) = 0$, it also follows that $\hat{\rho}_{AB} - \rho \xrightarrow{d} \tilde{X}'_{61} W_{11} \tilde{X}_{51} / \tilde{X}'_{61} W_{11} \tilde{X}_{61}$.

Finally, it is easily seen that if $T = 3$, then $\tilde{X}_5 = \tilde{X}_{51} = \tilde{X}_1 / \sqrt{\lambda}$ and $\tilde{X}_6 = \tilde{X}_{61} = \tilde{X}_2 / \sqrt{\lambda}$, where \tilde{X}_1 and \tilde{X}_2 are defined in theorem 2.

Part b): The proof of the first two results is similar to the proof of the last part of theorem 2 and the proof of part a) of this theorem:

Stationarity implies that for $\rho = 1 - \lambda N^{-1}$ $\lim_{N \rightarrow \infty} E(y_{i,t} \Delta y_{i,t}) = \lim_{N \rightarrow \infty} E(y_{i,2} \Delta y_{i,2}) = \sigma^2/2$ and $\lim_{N \rightarrow \infty} N^{-1} \text{Var}(y_{i,t} \Delta y_{i,t}) = \lim_{N \rightarrow \infty} N^{-1} \text{Var}(y_{i,2} \Delta y_{i,2}) = \sigma^4/(2\lambda)$. Moreover, if $s < t$, then for $\rho = 1 - \lambda N^{-1}$ $\lim_{N \rightarrow \infty} N^{-1} E(y_{i,s} \Delta y_{i,s} y_{i,t} \Delta y_{i,t}) = \lim_{N \rightarrow \infty} N^{-1} E(u_{i,s} \Delta u_{i,s} \times u_{i,t} \Delta u_{i,t}) = \lim_{N \rightarrow \infty} N^{-1} E(u_{i,s-1} \Delta u_{i,s} u_{i,t-1} \Delta u_{i,t}) = \lim_{N \rightarrow \infty} N^{-1} E(u_{i,s-1} \Delta u_{i,s} u_{i,s-1} \Delta u_{i,s+1}) = \lim_{N \rightarrow \infty} N^{-1} E\{u_{i,s-1}^2 [(\rho - 1)u_{s-1} + \varepsilon_s][(\rho - 1)(\rho u_{s-1} + \varepsilon_s)]\} = 0$.

Thus for $\rho = 1 - \lambda N^{-1}$ $\lim_{N \rightarrow \infty} E(Z_i^{L'} v_i v_i' Z_i^L) = \lim_{N \rightarrow \infty} \sigma^2 E(Z_i^{L'} Z_i^L) = \sigma^4 I$, $\lim_{N \rightarrow \infty} N^{-1} E(Z_i^{L'} y_{i,-1} y'_{i,-1} Z_i^L) = (\sigma^4/2\lambda) I$ and $\lim_{N \rightarrow \infty} N^{-1/2} E(Z_i^{L'} y_{i,-1} v_i' Z_i^L) = 0$ and hence $\begin{bmatrix} N^{-1/2} \sum_{i=1}^N Z_i^{L'} v_i \\ N^{-1} \sum_{i=1}^N Z_i^{L'} y_{i,-1} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \tilde{X}_{71} \\ \tilde{X}_{81} \end{bmatrix} \sim N \left[\begin{pmatrix} 0 \\ \frac{1}{2} \sigma^2 t \end{pmatrix}, \begin{pmatrix} \sigma^4 \\ 2\lambda \end{pmatrix} \begin{pmatrix} 2\lambda I & 0 \\ 0 & I \end{pmatrix} \right]$.

Note that $\tilde{\Sigma}_{78,22} = E(\tilde{X}_{71} \tilde{X}'_{81}) = 0$. Since $\lim_{N \rightarrow \infty} N^{-1/2} E(Z_i^{D'} \Delta v_i y'_{i,-1} Z_i^L) = 0$, we also have $\tilde{\Sigma}_{78,12} = E(\tilde{X}_{52} \tilde{X}'_{81}) = 0$. We conclude that $\tilde{\Sigma}_{78} = 0$. Furthermore it is easily verified that

$$\tilde{\Sigma}_{77} = \text{Var}(\tilde{X}_7) = \lim_{N \rightarrow \infty} K^{II} E(Z_i^{II'} v_i v_i' Z_i^{II}) K^{II'} = \sigma^4 K^{II} K^{II'}.$$

It follows in a straightforward manner from the above results that $\text{plim}_{N \rightarrow \infty} (\hat{\rho}_{Arbov} - \rho) = 0$ and $\sqrt{N}(\hat{\rho}_{Arbov} - \rho) \xrightarrow{d} \tilde{X}'_8 W \tilde{X}_7 / \tilde{X}'_8 W \tilde{X}_8$ with $\tilde{X}_7 = (\tilde{X}'_{52} \tilde{X}'_{71})' \sim N(0, \tilde{\Sigma}_{77})$ and $\tilde{X}_8 = (0' \tilde{X}'_{81})'$.

Finally, it is easily seen that if $T = 3$, then $\tilde{X}_7 = \tilde{X}_3$ and $\tilde{X}_8 = \tilde{X}_4$, where \tilde{X}_3 and \tilde{X}_4 are defined in theorem 2.

Part c): Since $\lim_{N \rightarrow \infty} N^{-2} \text{Var}(y_{i,1}^2) < \infty$, $\lim_{N \rightarrow \infty} N^{-1} \text{Var}(y_{i,1} \Delta y_{i,t}) < \infty$ and $\lim_{N \rightarrow \infty} \text{Var}([\Delta y_{i,t}]^2) < \infty$, it is easily seen that for $\rho = 1 - \lambda N^{-1}$, $N^{-2} \sum_{i=1}^N (Z_i^{II'} Z_i^I) \xrightarrow{q.m.} \lim_{N \rightarrow \infty} N^{-1} E(Z_i^{II'} Z_i^I) = \frac{\sigma^2}{2\lambda} I$, $N^{-3/2} \sum_{i=1}^N (Z_i^{II'} Z_i^D) \xrightarrow{q.m.} \lim_{N \rightarrow \infty} N^{-1/2} E(Z_i^{II'} Z_i^D) = 0$ and $N^{-1} \sum_{i=1}^N (Z_i^{D'} Z_i^D) \xrightarrow{q.m.} \lim_{N \rightarrow \infty} E(Z_i^{D'} Z_i^D) = \sigma^2 I$. The results are obtained by noting that convergence in quadratic mean implies convergence in probability.

Part d): Since $\lim_{N \rightarrow \infty} \text{Var}([\Delta y_{i,t}]^2) < \infty \forall t$, it is easily seen that for $\rho = 1 - \lambda N^{-1}$ $N^{-1} \sum_{i=1}^N (Z_i^{III'} Z_i^{II}) \xrightarrow{q.m.} \lim_{N \rightarrow \infty} E(Z_i^{III'} Z_i^{II}) = \sigma^2 I$ and hence $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N (Z_i^{III'} Z_i^{II}) = \sigma^2 I$.

To establish the second claim of part d), note that $\hat{v}_i = y_i - \hat{\rho}_1 y_{i,-1} = v_i + (\rho - \hat{\rho}_1) y_{i,-1}$ and $\Delta \hat{v}_i = \Delta y_i - \hat{\rho}_1 \Delta y_{i,-1} = \Delta v_i + (\rho - \hat{\rho}_1) \Delta y_{i,-1}$.

Lemma 10 implies that $\lim_{N \rightarrow \infty} E(\Delta u_{i,t})^8 < \infty$, $\lim_{N \rightarrow \infty} N^{-1} E[u_{i,s}^2 (\Delta u_{i,t})^6] < \infty$, $\lim_{N \rightarrow \infty} N^{-2} E[u_{i,s}^4 (\Delta u_{i,t})^4] < \infty$, $\lim_{N \rightarrow \infty} E[(\Delta u_{i,t})^4] < \infty$, $\lim_{N \rightarrow \infty} |E[u_{i,s} (\Delta u_{i,t})^3]| < \infty$ and $\lim_{N \rightarrow \infty} N^{-1} E[u_{i,s}^2 (\Delta u_{i,t})^2] < \infty \forall s, t$. It follows from the first result, our model assumptions and the Cauchy-Schwarz (CS) inequality that $\lim_{N \rightarrow \infty} \text{Var}([v_{i,s} \Delta y_{i,t}]^2) < \infty \forall s, t$ and hence $N^{-1} \sum_{i=1}^N \tilde{Z}_i^{III'} [\Delta v_i' v_i'] [\Delta v_i' v_i'] \tilde{Z}_i^{II} \xrightarrow{q.m.} \lim_{N \rightarrow \infty} K^{II} E(Z_i^{III'} v_i v_i' Z_i^{II}) K^{II'} = \sigma^4 K^{II} K^{II'}$. Similarly it follows that $N^{-2} \sum_{i=1}^N (Z_i^{L'} y_{i,-1} y_{i,-1}' Z_i^L) \xrightarrow{q.m.} \lim_{N \rightarrow \infty} N^{-1} E(Z_i^{L'} y_{i,-1} y_{i,-1}' Z_i^L) = (\sigma^4 / 2\lambda) I$, $N^{-2} \sum_{i=1}^N (Z_i^{D'} \Delta y_{i,-1} \Delta y_{i,-1}' Z_i^D) \xrightarrow{q.m.} 0$ and $N^{-2} \sum_{i=1}^N (Z_i^{L'} y_{i,-1} \Delta y_{i,-1}' Z_i^D) \xrightarrow{q.m.} 0$.

Lemma 10 and our model assumptions also imply that $\lim_{N \rightarrow \infty} |E[v_{i,k} u_{i,s} (\Delta u_{i,t})^2]| < \infty$ and $\lim_{N \rightarrow \infty} N^{-1} E[v_{i,k}^2 u_{i,s}^2 (\Delta u_{i,t})^4] < \infty \forall k, s, t$. It follows from these results, our model assumptions and the CS inequality that $N^{-3/2} \sum_{i=1}^N (Z_i^{L'} v_i y_{i,-1}' Z_i^L) \xrightarrow{q.m.} 0$, $N^{-3/2} \sum_{i=1}^N (Z_i^{L'} v_i \Delta y_{i,-1}' Z_i^D) \xrightarrow{q.m.} 0$, $N^{-3/2} \sum_{i=1}^N (Z_i^{L'} y_{i,-1} \Delta v_i' Z_i^D) \xrightarrow{q.m.} 0$ and $N^{-3/2} \sum_{i=1}^N (Z_i^{D'} \Delta y_{i,-1} \Delta v_i' Z_i^D) \xrightarrow{q.m.} 0$.

We conclude from the above results that $N^{-1} \sum_{i=1}^N \tilde{Z}_i^{III'} [\widehat{\Delta v}_i' \hat{v}_i'] [\widehat{\Delta v}_i' \hat{v}_i'] \tilde{Z}_i^{II} \xrightarrow{d} \sigma^4 K^{II} K^{II'} + (\sigma^4 / 2\lambda) \mathcal{K}^2 \text{diag}(O_{m-(T-2)}, I_{T-2})$.

Part e): The proof is very similar to part h) of the proof of theorem 1 apart from the first paragraph: Again recall that $\hat{\rho}_{AB1} = \{\sum_{t=2}^{T-1} [\tilde{y}'_{-1,t-1} Z_t (Z_t' Z_t)^{-1} Z_t' \tilde{y}_{-1,t-1}]\}^{-1} \times \sum_{t=2}^{T-1} [\tilde{y}'_{-1,t-1} Z_t (Z_t' Z_t)^{-1} Z_t' \tilde{y}_t]$. Since $\sum_{t=2}^{T-1} [\tilde{y}'_{-1,t-1} Z_t (Z_t' Z_t)^{-1} Z_t' \tilde{v}_t] = (\sum_{i=1}^N \Delta y'_{i,-1} Z_i^{AB}) \times [\sum_{i=1}^N (Z_i^{AB'} H Z_i^{AB})]^{-1} \sum_{i=1}^N Z_i^{AB'} \Delta v_i$ and $\sum_{t=2}^{T-1} [\tilde{y}'_{-1,t-1} Z_t (Z_t' Z_t)^{-1} Z_t' \tilde{y}_{-1,t-1}] = (\sum_{i=1}^N \Delta y'_{i,-1} Z_i^{AB}) [\sum_{i=1}^N (Z_i^{AB'} H Z_i^{AB})]^{-1} \sum_{i=1}^N Z_i^{AB'} \Delta y_{i,-1}$, it follows from part a) above that $\sum_{t=2}^{T-1} [\tilde{y}'_{-1,t-1} Z_t (Z_t' Z_t)^{-1} Z_t' \tilde{y}_t] \xrightarrow{d} \tilde{X}'_6 \tilde{\Sigma}_{55}^{-1} \tilde{X}_5$ and $\sum_{t=2}^{T-1} [\tilde{y}'_{-1,t-1} Z_t (Z_t' Z_t)^{-1} Z_t' \tilde{y}_{-1,t-1}] \xrightarrow{d} \tilde{X}'_6 \tilde{\Sigma}_{55}^{-1} \tilde{X}_5$. The rest of the proof is the same as part h) of the proof of theorem 1. \square

A.4 Proof of theorem 4

Part a): $W_{N,SY S1} = (N^{-1} \sum_{i=1}^N Z_i^{S'} A Z_i^S)^{-1}$ where $Z_i^S = \text{diag}(Z_i^I, Z_i^{II})$ and $A = \begin{bmatrix} H & C \\ C' & I_{T-2} \end{bmatrix}$.

Let $M_{11} = N^{-1} \sum_{i=1}^N (Z_i^{I'} H Z_i^I)$, $M_{12} = M'_{21} = N^{-1} \sum_{i=1}^N (Z_i^{I'} C Z_i^{II})$ and $M_{22} = N^{-1} \sum_{i=1}^N (Z_i^{II'} Z_i^{II})$. From parts a) and d) of the proof of theorem 3, we have $\text{plim}_{N \rightarrow \infty} N^{-1} \times M_{11} = \sigma^{-2} \tilde{\Sigma}_{55,11} = (\frac{\sigma^2}{2\lambda}) H$ and $\text{plim}_{N \rightarrow \infty} M_{22} = \sigma^2 I$. Moreover, since $\lim_{N \rightarrow \infty} E(y_{i,1} \Delta y_{i,t}) = -\sigma^2/2$ and $\lim_{N \rightarrow \infty} N^{-1} \text{Var}(y_{i,1} \Delta y_{i,t}) = \sigma^4/2\lambda$, we have $\lim_{N \rightarrow \infty} |E(Z_i^{I'} C Z_i^{II})| < \infty$ and $\text{plim}_{N \rightarrow \infty} N^{-1/2} M_{12} = \text{plim}_{N \rightarrow \infty} N^{-1/2} M'_{21} = 0$.

Note that $W_{N,SY S1} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}^{-1} = \begin{bmatrix} M^{11} & M^{12} \\ M^{21} & M^{22} \end{bmatrix}$, where

$M^{11} = N^{-1} (M_{11}/N - M_{12} M_{22}^{-1} M_{21}/N)^{-1}$, $M^{12} = -M^{11} M_{12} M_{22}^{-1}$, $M^{21} = -M_{22}^{-1} M_{21} M^{11}$, and $M^{22} = M_{22}^{-1} + M_{22}^{-1} M_{21} M^{11} M_{12} M_{22}^{-1}$.

The first claim follows now straightforwardly by Slutsky's theorem. The proof of the second claim is very similar.

Part b): Note that $\tilde{Z}_i^{S'} [\widehat{\Delta v}_i' \widehat{v}_i']' = ((Z_i^{I'} \widehat{\Delta v}_i)' (\tilde{Z}_i^{II'} [\widehat{\Delta v}_i' \widehat{v}_i']'))'$.

From part d) of theorem 3 we have $N^{-1} \sum_{i=1}^N \tilde{Z}_i^{II'} [\widehat{\Delta v}_i' \widehat{v}_i']' [\widehat{\Delta v}_i' \widehat{v}_i] \tilde{Z}_i^{II} \xrightarrow{d} \sigma^4 \tilde{W}_{Arbov2}^{-1}$.

Lemma 10, our model assumptions and the Cauchy-Schwarz inequality imply that $\lim_{N \rightarrow \infty} N^{-1} E[(\Delta v_{i,s})^2 u_{i,t}^2] < \infty$ and $\lim_{N \rightarrow \infty} N^{-2} E[(\Delta v_{i,s})^4 u_{i,t}^4] < \infty \quad \forall s, t$, and hence $N^{-2} \sum_{i=1}^N (Z_i^{II'} \Delta v_i \Delta v_i' Z_i^I) \xrightarrow{q.m.} \lim_{N \rightarrow \infty} N^{-1} E(Z_i^{II'} \Delta v_i \Delta v_i' Z_i^I)$. In part a) of the proof of theorem 3 we showed that $\lim_{N \rightarrow \infty} N^{-1} E(Z_i^{II'} \Delta v_i \Delta v_i' Z_i^I) = \lim_{N \rightarrow \infty} \sigma^2 N^{-1} E(Z_i^{I'} H Z_i^I) = (\frac{\sigma^4}{2\lambda}) H$. It then follows along the lines of part d) of the proof of theorem 3 that $\text{plim}_{N \rightarrow \infty} N^{-2} \sum_{i=1}^N (Z_i^{II'} \widehat{\Delta v}_i \widehat{\Delta v}_i' Z_i^I) = (\frac{\sigma^4}{2\lambda}) H$.

Similar arguments show that $\text{plim}_{N \rightarrow \infty} N^{-3/2} \sum_{i=1}^N (K^{II} Z_i^{II'} v_i \Delta v_i' Z_i^I) = 0$ and $\text{plim}_{N \rightarrow \infty} N^{-2} \sum_{i=1}^N (K^{II} Z_i^{II'} y_{i,-1} \Delta v_i' Z_i^I) = \text{plim}_{N \rightarrow \infty} N^{-2} \sum_{i=1}^N (y_{i,-1}^2 K^{II} Z_i^{II'} \iota \Delta v_i') = \frac{\sigma^2}{2\lambda} K^{II} \times E[\text{diag}(\varepsilon_i^2, \dots, \varepsilon_i^{T-1})' \iota \Delta \varepsilon_i'] = \frac{\sigma^4}{2\lambda} K^{II} D$. Note that $K^{II} Z_i^{II'} \widehat{v}_i = \widetilde{Z}_i^{II'} [\widehat{\Delta v}_i' \widehat{v}_i']$. It follows that $N^{-3/2} \sum_{i=1}^N (\widetilde{Z}_i^{II'} [\widehat{\Delta v}_i' \widehat{v}_i'] \widehat{\Delta v}_i' Z_i^I) \xrightarrow{d} -\frac{\sigma^4}{2\lambda} \mathcal{K} K^{II} D$.

Noting that $\widetilde{W}_{SY S_2}^{11} = \frac{1}{2\lambda} H$, $\widetilde{W}_{SY S_2}^{21} = (\widetilde{W}_{SY S_2}^{12})' = -\frac{1}{2\lambda} \mathcal{K} K^{II} D$, and $\widetilde{W}_{SY S_2}^{22} = \widetilde{W}_{Arbov_2}^{-1}$, we conclude that $\text{diag}(N^{-1} I_{T-2}, N^{-1/2} I_m) \sum_{i=1}^N \widetilde{Z}_i^{S'} [\widehat{\Delta v}_i' \widehat{v}_i'] [\widehat{\Delta v}_i' \widehat{v}_i'] \widetilde{Z}_i^S \text{diag}(N^{-1} I_{T-2}, N^{-1/2} I_m) \xrightarrow{d} \sigma^4 \widetilde{W}_{SY S_2}^{-1}$.

Part c): From part a) of theorem 3 we have $N^{-1} \sum_{i=1}^N Z_i^{II'} \Delta v_i \xrightarrow{d} \widetilde{X}_{51}$, $N^{-1} \sum_{i=1}^N Z_i^{II'} \Delta y_{i,-1} \xrightarrow{d} \widetilde{X}_{61}$, $N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta v_i \xrightarrow{d} \widetilde{X}_{52}$, and $N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta y_{i,-1} \xrightarrow{d} \widetilde{X}_{62}$, with $\widetilde{X}_5 = (\widetilde{X}_{51}' \ \widetilde{X}_{52}')' \sim N(0, \widetilde{\Sigma}_{55})$ and $\widetilde{X}_6 = (\widetilde{X}_{61}' \ \widetilde{X}_{62}')' \sim N(\widetilde{\mu}_6, \widetilde{\Sigma}_{66})$.

From part b) of theorem 3 we have $N^{-1/2} \sum_{i=1}^N Z_i^{L'} v_i \xrightarrow{d} \widetilde{X}_{71} \sim N(0, \widetilde{\Sigma}_{77,22})$, and $N^{-1} \sum_{i=1}^N Z_i^{L'} y_{i,-1} \xrightarrow{d} \widetilde{X}_{81} \sim N(\frac{\sigma^2 \iota}{2}, \frac{\sigma^4}{2\lambda} I)$.

Recall that $\widetilde{Z}_i^{II'} [\Delta v_i' \ v_i']' = K^{II} Z_i^{II'} v_i$. Let $K^S = \text{diag}(I_{T-2}, K^{II})$. It follows that $\widetilde{Z}_i^{S'} [\Delta v_i' \ v_i']' = ((Z_i^{I'} \widehat{\Delta v}_i)' (\widetilde{Z}_i^{II'} [\widehat{\Delta v}_i' \widehat{v}_i']'))' = K^S Z_i^{S'} [\Delta v_i' \ v_i']'$.

Consider now $K^S W_{N, SY S_1}^{-1} K^{S'} = N^{-1} \sum_{i=1}^N K^S Z_i^{S'} A Z_i^S K^{S'}$ and $K^S W_{N, SY S_{1b}}^{-1} K^{S'} = N^{-1} \sum_{i=1}^N K^S Z_i^{S'} Z_i^S K^{S'}$. From part a) we obtain $\text{plim}_{N \rightarrow \infty} N^{-2} \sum_{i=1}^N (Z_i^{II'} H Z_i^I) = (\frac{\sigma^2}{2\lambda}) H$, $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N (K^{II} Z_i^{II'} Z_i^{II} K^{II'}) = \sigma^2 K^{II} K^{II'}$ and $\sigma^{-2} \widetilde{\Sigma}_{77}$, and $\text{plim}_{N \rightarrow \infty} N^{-3/2} \sum_{i=1}^N (Z_i^{I'} C Z_i^{II} K^{II'}) = 0$. We also have $\text{plim}_{N \rightarrow \infty} N^{-2} \sum_{i=1}^N (Z_i^{II'} Z_i^I) = (\frac{\sigma^2}{2\lambda}) I$.

We conclude from the above results that $\text{plim}_{N \rightarrow \infty} (\widehat{\rho}_{SY S_k} - \rho) = 0$ and $\sqrt{N}(\widehat{\rho}_{SY S_k} - \rho) \xrightarrow{d} \widetilde{X}_8' \widetilde{\Sigma}_{77}^{-1} \widetilde{X}_7 / \widetilde{X}_8' \widetilde{\Sigma}_{77}^{-1} \widetilde{X}_8$, for $k = 1, 1b$.

It immediately follows from part b) and the above results that $\text{plim}_{N \rightarrow \infty} (\widehat{\rho}_{SY S_2} - \rho) = 0$ and $\sqrt{N}(\widehat{\rho}_{SY S_2} - \rho) \xrightarrow{d} \widetilde{X}_{11}' \widetilde{W}_{SY S_2} \widetilde{X}_{10} / \widetilde{X}_{11}' \widetilde{W}_{SY S_2} \widetilde{X}_{11}$ with $\widetilde{X}_{10} = (\widetilde{X}_{51}' \ \widetilde{X}_{52}' \ \widetilde{X}_{71}')'$ and $\widetilde{X}_{11} = (0 \ 0 \ \widetilde{X}_{81}')'$.

Part d): From the results mentioned in part c) above we conclude $\widehat{\rho}_{SY S} - \rho \xrightarrow{d} (\widetilde{X}_{61}' W_{11} \widetilde{X}_{51} + \widetilde{X}_{81}' W_{31} \widetilde{X}_{51}) / \widetilde{X}_{12}' W \widetilde{X}_{12}$ with $\widetilde{X}_{12} = (\widetilde{X}_{61}' \ 0 \ \widetilde{X}_{81}')'$. \square

Lemma 11 *Let g, d, \hat{c} and λ be constants such that $0 < g \leq 1$, $0 \leq d \leq g$, $0 < \hat{c} < \infty$, and $0 < \lambda < \infty$. Furthermore, let $\rho = 1 - \lambda N^{-g}$. Then $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{g-d} (1 - \rho^{2(1+S)}) \equiv \hat{q}(\lambda, \hat{c}) < \infty$, where $\hat{q}(\lambda, \hat{c}) = 2\lambda(S+1)$ if $d = 0$; $\hat{q}(\lambda, \hat{c}) = 2\lambda\hat{c}$ if $0 < d < g$; and $\hat{q}(\lambda, \hat{c}) = q(\lambda, \hat{c})$ if $d = g$.*

Proof of lemma 11: Note that $(1 - \rho^{2(1+S)})/(1 - \rho) = (1 + \rho^{1+S}) \sum_{k=0}^S \rho^k$ and recall that $\lim_{p \rightarrow \infty} (1 + x/p)^p = \exp(x)$ and $q(\lambda, c) = 1 - \exp(-2\lambda c)$. There are three cases: $d = 0$, $0 < d < g$, and $d = g$.

First assume that $d = 0$. Then $S = \hat{c}$ is fixed and $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{g-d} (1 - \rho^{2(1+S)}) = \lim_{N \rightarrow \infty, S=\hat{c}} \lambda (1 - \rho^{2(1+S)}) / (1 - \rho) = \lim_{N \rightarrow \infty, S=\hat{c}} \lambda (1 + \rho^{1+S}) \sum_{k=0}^S \rho^k = 2\lambda(S + 1)$.

Next assume that $0 < d < g$. Then we have $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} S^{-1} \sum_{k=0}^S \rho^k = 1$, because $1 \geq \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} S^{-1} \sum_{k=0}^S \rho^k \geq \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} \rho^S = 1$. It follows that $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{g-d} (1 - \rho^{2(1+S)}) = \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} \lambda N^{-d} (1 - \rho^{2(1+S)}) / (1 - \rho) = \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} \lambda (1 + \rho^{1+S}) \hat{c} S^{-1} \sum_{k=0}^S \rho^k = 2\lambda \hat{c}$.

Finally assume that $d = g$. Then we have $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{g-d} (1 - \rho^{2(1+S)}) = \lim_{N \rightarrow \infty, S/N^g \rightarrow \hat{c}} (1 - \rho^{2(1+S)}) = 1 - \exp(-2\lambda \hat{c}) = q(\lambda, \hat{c}) < \infty$. \square

A.5 Proof of theorem 5

Note that $\lim_{\rho \uparrow 1} \text{Var}(y_{i,-S} - \mu_i) = 0$. Then we obtain for the parameter sequence $\rho = 1 - \lambda N^{-g}$ that $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-d} \text{Var}(y_{i,1} - \mu_i) = \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} \sigma^2 N^{-d} (1 - \rho^{2(1+S)}) / (1 - \rho^2) = \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} \sigma^2 N^{g-d} (1 - \rho^{2(1+S)}) / (2\lambda) = \sigma^2 \hat{q}(\lambda, \hat{c}) / (2\lambda)$ by lemma 11.

Let $\sigma_{-S}^2 = E(y_{i,-S}^2)$ and let the indicator function $1\{d = 0\} = 1$ if $d = 0$ and let $1\{d = 0\} = 0$ if $d \neq 0$. Then it follows that $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-d} \text{Var}(y_{i,1} \Delta y_{i,2}) = \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-d} \text{Var}(y_{i,1} \varepsilon_{i,2}) = \sigma^2 (\sigma^2 \hat{q}(\lambda, \hat{c}) / (2\lambda) + 1\{d = 0\} \sigma_{-S}^2)$. Furthermore, if $0 < d \leq g$, we have $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{g-d} E(y_{i,1} \Delta y_{i,2}) = -\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} \lambda N^{-d} \text{Var}(y_{i,1} - \mu_i) = -\sigma^2 \hat{q}(\lambda, \hat{c}) / 2$. Finally, if $d = 0$, we obtain by using arguments similar to those in the proof of theorem 1 that $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{g-d} E(y_{i,1} \Delta y_{i,2}) = \lambda (\sigma_{\mu y} - \sigma_y^2)$. Note that when $d = 0$, $\sigma_y^2 = \sigma^2 \hat{q}(\lambda, \hat{c}) / (2\lambda) + \sigma_{-S}^2$.

When $T = 3$, $\hat{\rho}_{AB} = \rho + \frac{\sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3}}{\sum_{i=1}^N y_{i,1} \Delta y_{i,2}}$ and $\hat{\rho}_{Arbov} = \rho + \frac{\sum_{i=1}^N (\varepsilon_{i,3} + (1-\rho)\mu_i) \Delta y_{i,2}}{\sum_{i=1}^N y_{i,2} \Delta y_{i,2}}$.

Let us define $X_1 = N^{-\frac{1}{2}(1+d)} \sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3}$, $X_3 = N^{-1/2} \sum_{i=1}^N (\varepsilon_{i,3} + (1-\rho)\mu_i) \Delta y_{i,2}$ and $X_4 = N^{-1} \sum_{i=1}^N y_{i,2} \Delta y_{i,2}$. Moreover, let $X_2 = N^{-\frac{1}{2}(1+d)} \sum_{i=1}^N y_{i,1} \Delta y_{i,2}$ if $\Delta_{gd} \leq 0$ and $X_2 = N^{-1+g-d} \sum_{i=1}^N y_{i,1} \Delta y_{i,2}$ if $\Delta_{gd} > 0$.

Recall that $q(\lambda, c) = 0$ if $c = 0$. Then using results similar to those in lemma 9 we obtain as $N, S \rightarrow \infty$ with $S/N \rightarrow c \geq 0$, $S/N^g \rightarrow \bar{c} \geq 0$ and $S/N^d \rightarrow \hat{c} > 0$, where $0 < g \leq 1$ and $0 \leq d \leq g$, for the parameter sequence $\rho = 1 - \lambda N^{-g}$ that $X_1 \xrightarrow{d} N(0, 2\sigma^2 (\sigma^2 \hat{q}(\lambda, \hat{c}) / (2\lambda) +$

$1\{d = 0\}\sigma_{-S}^2$), $X_3 \xrightarrow{d} N(0, \sigma^4)$ and $X_4 \xrightarrow{d} N(\sigma^2(1 - q(\lambda, \bar{c})/2), \sigma^4 q(\lambda, c)/(2\lambda))$. In addition, we obtain that

$$\begin{aligned} X_2 &\xrightarrow{d} N(\lambda(\sigma_{\mu y} - \sigma_y^2), \sigma^2(\sigma^2 \hat{q}(\lambda, \hat{c})/(2\lambda) + \sigma_{-S}^2)) \text{ if } \Delta_{gd} = 0 \text{ and } d = 0, \\ X_2 &\xrightarrow{d} N(-\sigma^2 \hat{q}(\lambda, \hat{c})/2, \sigma^4 \hat{q}(\lambda, \hat{c})/(2\lambda)) \text{ if } \Delta_{gd} = 0 \text{ and } d > 0, \\ X_2 &\xrightarrow{d} N(0, \sigma^2(\sigma^2 \hat{q}(\lambda, \hat{c})/(2\lambda) + 1\{d = 0\}\sigma_{-S}^2)) \text{ if } \Delta_{gd} < 0, \\ X_2 &\xrightarrow{d} \lambda(\sigma_{\mu y} - \sigma_y^2) \text{ if } \Delta_{gd} > 0 \text{ and } d = 0, \text{ and} \\ X_2 &\xrightarrow{d} -\sigma^2 \hat{q}(\lambda, \hat{c})/2 \text{ if } \Delta_{gd} > 0 \text{ and } d > 0, \end{aligned}$$

and we obtain that $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} Cov(X_1, X_2) = -\sigma^2(\sigma^2 \hat{q}(\lambda, \hat{c})/(2\lambda) + 1\{d = 0\}\sigma_{-S}^2)$ if $\Delta_{gd} \leq 0$, $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} Cov(X_1, X_2) = 0$ if $\Delta_{gd} > 0$, and $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} Cov(X_3, X_4) = 0$.

Finally note that if $\Delta_{gd} = 0$ and $d = 0$, then $g = 1/2$. This case corresponds to local-to-zero asymptotics, see theorem 1. Parts a) and b) of theorem 5 follow now straightforwardly from the above results. \square

A.6 Proof of theorem 6

When $T = 3$ and $\rho = 1$, $\hat{\rho}_{AB} = \rho + \frac{\sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3}}{\sum_{i=1}^N y_{i,1} \varepsilon_{i,2}}$ and $\hat{\rho}_{Arbov} = \rho + \frac{\sum_{i=1}^N \varepsilon_{i,3} \varepsilon_{i,2}}{\sum_{i=1}^N y_{i,2} \varepsilon_{i,2}}$.

Let us define $X_1 = N^{-\frac{1}{2}(1+d)} \sum_{i=1}^N y_{i,1} \Delta \varepsilon_{i,3}$, $X_2 = N^{-\frac{1}{2}(1+d)} \sum_{i=1}^N y_{i,1} \varepsilon_{i,2}$, $X_3 = N^{-1/2} \sum_{i=1}^N \varepsilon_{i,3} \varepsilon_{i,2}$ and $X_4 = N^{-\frac{1}{2}(1+\bar{d})} \sum_{i=1}^N y_{i,2} \varepsilon_{i,2}$. Let $\sigma_{-S}^2 = E(y_{i,-S}^2) + \sigma^2$. Finally, let the indicator function $1\{d = 0\} = 1$ if $d = 0$ and let $1\{d = 0\} = 0$ if $d \neq 0$. Then it is easily seen that as $N, S \rightarrow \infty$ with $S/N^d \rightarrow \hat{c} > 0$ and $S/N^{\bar{d}} \rightarrow c \geq 0$ where $d \geq 0$ and $\bar{d} = \max(d, 1)$, one obtains that $X_1 \xrightarrow{d} N(0, 2(\hat{c}\sigma^2 + 1\{d = 0\}\sigma_{-S}^2)\sigma^2)$, $X_2 \xrightarrow{d} N(0, (\hat{c}\sigma^2 + 1\{d = 0\}\sigma_{-S}^2)\sigma^2)$, $X_3 \xrightarrow{d} N(0, \sigma^4)$ and $X_4 \xrightarrow{d} N(1\{d \leq 1\}\sigma^2, c\sigma^4)$. In addition, one obtains that $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} Cov(X_1, X_2) = -(\hat{c}\sigma^2 + 1\{d = 0\}\sigma_{-S}^2)\sigma^2$ and $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} Cov(X_3, X_4) = 0$. Parts a) and b) of theorem 6 follow now straightforwardly from the above results. \square

A.7 Proof of theorem 7

Part a): Recall that $\tilde{Z}_i^{II} = \text{diag}(Z_i^D, Z_i^L)$ and that $\tilde{Z}_i^S = \text{diag}(Z_i^{AB}, Z_i^L)$, where $Z_i^{AB} = [Z_i^I \ Z_i^D]$. Also recall that $K^{II} Z_i^{II'} v_i = \tilde{Z}_i^{II'} [\Delta v_i' \ v_i']'$ and from the proof of theorem 4 that $K^S Z_i^{S'} [\Delta v_i' \ v_i']' = \tilde{Z}_i^{S'} [\Delta v_i' \ v_i']'$, when $K^S = \text{diag}(I_{T-2}, K^{II})$. Similarly, we have $K^{II} Z_i^{II'} y_{i,-1} = \tilde{Z}_i^{II'} [\Delta y_{i,-1}' \ y_{i,-1}']'$ and $K^S Z_i^{S'} [\Delta y_{i,-1}' \ y_{i,-1}']' = \tilde{Z}_i^{S'} [\Delta y_{i,-1}' \ y_{i,-1}']'$.

Note that when $\rho = 1$, $v_{i,t} = \varepsilon_{i,t}$, $\Delta y_{i,t} = \varepsilon_{i,t}$, $\Delta y_i^t = \varepsilon_i^t$ and $Z_i^{II} = \text{diag}(\varepsilon_i^2, \dots, \varepsilon_i^{T-1})$. Now let $\sigma_{-S}^2 = E(y_{i,-S}^2) + \sigma^2$. Furthermore, let the indicator function $1\{d = 0\} = 1$ if $d = 0$ and let $1\{d = 0\} = 0$ if $d \neq 0$. Similarly let $1\{d \leq 1\} = 1$ if $d \leq 1$ and let $1\{d \leq 1\} = 0$ if $d > 1$. Then as $N, S \rightarrow \infty$ with $S/N^d \rightarrow \hat{c} > 0$ and $S/N^{\bar{d}} \rightarrow c \geq 0$ where $d \geq 0$ and $\bar{d} = \max(d, 1)$, we obtain that

$$\begin{aligned} N^{-\frac{1}{2}(1+d)} \sum_{i=1}^N Z_i^{II} \Delta v_i &\xrightarrow{d} \hat{X}_{51} \sim N(0, (\hat{c}\sigma^2 + 1\{d = 0\}\sigma_{-S}^2)\sigma^2 H), \\ N^{-\frac{1}{2}(1+d)} \sum_{i=1}^N Z_i^{II} \Delta y_{i,-1} &\xrightarrow{d} \hat{X}_{61} \sim N(0, (\hat{c}\sigma^2 + 1\{d = 0\}\sigma_{-S}^2)\sigma^2 I), \\ N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta v_i &\xrightarrow{d} \hat{X}_{52} \sim N(0, \sigma^4 (H_{T-3}, H_{T-4}, \dots, H_1)), \\ N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta y_{i,-1} &\xrightarrow{d} \hat{X}_{62} \sim N(0, \sigma^4 I), \\ N^{-1/2} \sum_{i=1}^N Z_i^{L'} v_i &\xrightarrow{d} \hat{X}_{71} \sim N(0, \sigma^4 I), \\ N^{-\frac{1}{2}(1+\bar{d})} \sum_{i=1}^N Z_i^{L'} y_{i,-1} &\xrightarrow{d} \hat{X}_{81} \sim N(1\{d \leq 1\}\sigma^2 \iota, c\sigma^4 I), \\ \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} \text{Cov}(N^{-\frac{1}{2}(1+\bar{d})} \sum_{i=1}^N Z_i^{L'} y_{i,-1}, N^{-1/2} \sum_{i=1}^N Z_i^{D'} \Delta v_i) &= 0, \text{ and} \\ \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} \text{Cov}(N^{-\frac{1}{2}(1+\bar{d})} \sum_{i=1}^N Z_i^{L'} y_{i,-1}, N^{-1/2} \sum_{i=1}^N Z_i^{L'} v_i) &= 0. \end{aligned}$$

Let $\hat{X}_7 = (\hat{X}'_{52} \hat{X}'_{71})'$ and $\hat{X}_8 = (0' \hat{X}'_{81})'$. From the above results we have $\text{Cov}(\hat{X}_7, \hat{X}_8) = E(\hat{X}_7 \hat{X}_8') = 0$ and $\hat{X}_7 \perp \hat{X}_8$.

Note that $W_{N,Arbov2}(1) = (N^{-1} \sum_{i=1}^N Z_i^{II'} v_i v_i' Z_i^{II})^{-1}$ and $W_{N,SYs2}(1) = (N^{-1} \sum_{i=1}^N Z_i^{S'} \times [\Delta v_i' v_i']' [\Delta v_i' v_i'] Z_i^S)^{-1}$. Consider now $K^{II} W_{N,Arbov2}^{-1}(1) K^{II'}$ and $K^S W_{N,SYs2}^{-1}(1) K^{S'}$. Recall that $\sigma^4 K^{II} K^{II'} = \tilde{\Sigma}_{77}$. Then it is easy to verify that

$$\begin{aligned} \text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-(1+d)} \sum_{i=1}^N Z_i^{II'} \Delta v_i \Delta v_i' Z_i^{II} &= (\hat{c}\sigma^2 + 1\{d = 0\}\sigma_{-S}^2)\sigma^2 H, \\ \text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-1} \sum_{i=1}^N K^{II} Z_i^{II'} v_i v_i' Z_i^{II} K^{II'} &= \tilde{\Sigma}_{77}, \text{ and} \\ \text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-\frac{1}{2}(2+d)} \sum_{i=1}^N K^{II} Z_i^{II'} v_i \Delta v_i' Z_i^{II} &= 0. \end{aligned}$$

We conclude from the above results that $\text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} (\hat{\rho}_{Arbov2,R} - 1) = 0$ and $N^{\frac{1}{2}\bar{d}} (\hat{\rho}_{Arbov2,R} - 1) \xrightarrow{d} \hat{X}_8' \tilde{\Sigma}_{77}^{-1} \hat{X}_7 / \hat{X}_8' \tilde{\Sigma}_{77}^{-1} \hat{X}_8$. Furthermore, we can conclude that $\text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} (\hat{\rho}_{SYs2,R} - 1) = 0$ and $N^{\frac{1}{2}\bar{d}} (\hat{\rho}_{SYs2,R} - 1) \xrightarrow{d} \hat{X}_8' \tilde{\Sigma}_{77}^{-1} \hat{X}_7 / \hat{X}_8' \tilde{\Sigma}_{77}^{-1} \hat{X}_8$.

Note that $SE(\hat{\rho}_{Arbov2,\rho}) = [N^{-1} \sum_{i=1}^N (y'_{i,-1} Z_i^{II'} K^{II'}) (K^{II'})^{-1} W_{N,Arbov2}(1) (K^{II})^{-1} \times \sum_{i=1}^N (K^{II} Z_i^{II'} y_{i,-1})]^{-1/2}$ and that $SE(\hat{\rho}_{SYs2,R}) = [N^{-1} \sum_{i=1}^N ((\Delta y'_{i,-1} y'_{i,-1}) Z_i^S K^{S'}) \times (K^{S'})^{-1} W_{N,SYs2}(1) (K^S)^{-1} \sum_{i=1}^N (K^S Z_i^{S'} (\Delta y'_{i,-1} y'_{i,-1}))]^{-1/2}$.

It follows from the above results that $[SE(\hat{\rho}_{Arbov2,R})]^{-2}/N^{\bar{d}} \xrightarrow{d} \hat{X}_8' \tilde{\Sigma}_{77}^{-1} \hat{X}_8$ and hence $(\hat{\rho}_{Arbov2,R} - 1)/SE(\hat{\rho}_{Arbov2,R}) \xrightarrow{d} (\hat{X}_8' \tilde{\Sigma}_{77}^{-1} \hat{X}_8)^{-1/2} \hat{X}_8' \tilde{\Sigma}_{77}^{-1} \hat{X}_7$. Since $\hat{X}_7 \perp \hat{X}_8$, $(\hat{X}_8' \tilde{\Sigma}_{77}^{-1} \hat{X}_8)^{-1/2} \times \hat{X}_8' \tilde{\Sigma}_{77}^{-1} \hat{X}_7 | \hat{X}_8 \sim N(0, 1)$ and therefore $(\hat{\rho}_{Arbov2,R} - 1)/SE(\hat{\rho}_{Arbov2,R}) \xrightarrow{d} N(0, 1)$. Similarly we obtain that $[SE(\hat{\rho}_{SYs2,R})]^{-2}/N^{\bar{d}} \xrightarrow{d} \hat{X}_8' \tilde{\Sigma}_{77}^{-1} \hat{X}_8$ and $(\hat{\rho}_{SYs2,R} - 1)/SE(\hat{\rho}_{SYs2,R}) \xrightarrow{d} N(0, 1)$.

Part b): We first consider the weight matrix and some other fourth order moments. Note that $\rho = 1 - \lambda N^{-1/2}$. Let $u_{i,t} = y_{i,t} - \mu_i \forall t$. Then $\lim_{S,N \rightarrow \infty, seq} E[(\Delta u_{i,t})^8] < \infty$, $\lim_{S,N \rightarrow \infty, seq} E[(\Delta u_{i,t})^2] = \sigma^2$ and provided $s \neq t$, $\lim_{S,N \rightarrow \infty, seq} E(\Delta u_{i,s} \Delta u_{i,t}) = 0$. It follows from these results, our model assumptions and the Cauchy-Schwarz inequality that $\lim_{S,N \rightarrow \infty, seq} E(Z_i^{II'} v_i v_i' Z_i^{II}) = \sigma^4 I$ and $\lim_{S,N \rightarrow \infty, seq} Var([v_{i,s} \Delta y_{i,t}]^2) < \infty \forall s, t$. Consequently we obtain that $N^{-1} \sum_{i=1}^N Z_i^{II'} v_i v_i' Z_i^{II} \xrightarrow{q.m.} \sigma^4 I$ as $(S, N \rightarrow \infty)_{seq}$ and hence $\text{plim}_{S,N \rightarrow \infty, seq} N^{-1} \sum_{i=1}^N Z_i^{II'} v_i v_i' Z_i^{II} = \sigma^4 I$. Similar arguments show that $\text{plim}_{S,N \rightarrow \infty, seq} N^{-1} \sum_{i=1}^N Z_i^{II'} \Delta y_i \Delta y_i' Z_i^{II} = \sigma^4 I$, $\text{plim}_{S,N \rightarrow \infty, seq} N^{-3/2} \sum_{i=1}^N Z_i^{II'} \times (\Delta y_i - \Delta y_{i-1})(\Delta y_i - \Delta y_{i-1})' Z_i^{II} = \text{plim}_{S,N \rightarrow \infty, seq} N^{-3/2} \sum_{i=1}^N Z_i^{II'} \Delta v_i \Delta v_i' Z_i^{II} = (\frac{\sigma^4}{2\lambda})H$, and $\text{plim}_{S,N \rightarrow \infty, seq} N^{-5/4} \sum_{i=1}^N Z_i^{II'} (\Delta y_i - \Delta y_{i-1}) \Delta y_i' Z_i^{II} = \text{plim}_{S,N \rightarrow \infty, seq} N^{-5/4} \sum_{i=1}^N (Z_i^{II'} \Delta v_i \times v_i' Z_i^{II}) = 0$.

Next we consider second moments. By lemma 9 we have that $\lim_{S,N \rightarrow \infty, seq} E(u_{i,t} \Delta u_{i,t}) = \sigma^2/2 \forall t$, while lemma 10 yields $\lim_{S,N \rightarrow \infty, seq} N^{-1} E(u_{i,t}^2 (\Delta u_{i,t})^2) = 0 \forall t$. It follows that $N^{-1} \sum_{i=1}^N Z_i^{II'} y_{i,-1} \xrightarrow{q.m.} (\sigma^2/2)\iota$ as $(S, N \rightarrow \infty)_{seq}$ and hence $\text{plim}_{S,N \rightarrow \infty, seq} N^{-1} \sum_{i=1}^N Z_i^{II'} y_{i,-1} \equiv \vec{X}_8 = (\sigma^2/2)\iota$. Similarly, we obtain that $\text{plim}_{S,N \rightarrow \infty, seq} N^{-1} \sum_{i=1}^N Z_i^{II'} \Delta y_{i,-1} = -(\sigma^2/2)\iota$.

The above results imply that both $\text{plim}_{S,N \rightarrow \infty, seq} [SE(\hat{\rho}_{Arbov2,\rho})]^{-2}/N = \vec{X}_8' \vec{X}_8 / \sigma^4 = (1/2)^2 m$ and $\text{plim}_{S,N \rightarrow \infty, seq} [SE(\hat{\rho}_{SYS2,R})]^{-2}/N = \vec{X}_8' \vec{X}_8 / \sigma^4$.

Since $N^{-3/4} \sum_{i=1}^N Z_i^{II'} \Delta v_i \xrightarrow{d} \vec{X}_{51} \sim N(0, (\frac{\sigma^4}{2\lambda})H)$ and $N^{-1/2} \sum_{i=1}^N Z_i^{II'} v_i \xrightarrow{d} \vec{X}_7 \sim N(0, \sigma^4 I)$ as $(S, N \rightarrow \infty)_{seq}$, we also obtain that $(\hat{\rho}_{Arbov2,R} - \rho)/SE(\hat{\rho}_{Arbov2,R}) \xrightarrow{d} \sigma^{-2} (\vec{X}_8' \vec{X}_8)^{-1/2} \vec{X}_8' \vec{X}_7$ and $(\hat{\rho}_{SYS2,R} - \rho)/SE(\hat{\rho}_{SYS2,R}) \xrightarrow{d} \sigma^{-2} (\vec{X}_8' \vec{X}_8)^{-1/2} \vec{X}_8' \vec{X}_7$ as $(S, N \rightarrow \infty)_{seq}$.

Note that $\vec{X}_7 \perp \vec{X}_8$. Therefore $\sigma^{-2} (\vec{X}_8' \vec{X}_8)^{-1/2} \vec{X}_8' \vec{X}_7 | \vec{X}_8 \sim N(0, 1)$. We conclude that $(\hat{\rho}_{Arbov2,R} - 1)/SE(\hat{\rho}_{Arbov2,R}) \xrightarrow{d} N(-(\lambda/2)\sqrt{m}, 1)$ and $(\hat{\rho}_{SYS2,R} - 1)/SE(\hat{\rho}_{SYS2,R}) \xrightarrow{d} N(-(\lambda/2)\sqrt{m}, 1)$.

Part c): We first consider the weight matrix and some other fourth order moments. Note that $\rho = 1 - \lambda N^{-1/2}$ and $d \geq 0$. Let $u_{i,t} = y_{i,t} - \mu_i \forall t$. Then $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} E[(\Delta u_{i,t})^8] < \infty$, $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} E[(\Delta u_{i,t})^2] = \sigma^2$ and provided $s \neq t$, $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} E(\Delta u_{i,s} \Delta u_{i,t}) = 0$. It follows from these results, our model assumptions and the Cauchy-Schwarz inequality that $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} E(Z_i^{II'} v_i v_i' Z_i^{II}) = \sigma^4 I$ and $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} Var([v_{i,s} \Delta y_{i,t}]^2) < \infty \forall s, t$. Consequently we obtain that $N^{-1} \sum_{i=1}^N (Z_i^{II'} v_i \times v_i' Z_i^{II}) \xrightarrow{q.m.} \sigma^4 I$ and hence $\text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-1} \sum_{i=1}^N Z_i^{II'} v_i v_i' Z_i^{II} = \sigma^4 I$. Similar arguments

show that $\text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-1} \sum_{i=1}^N Z_i^{II'} \Delta y_i \Delta y_i' Z_i^{II} = \sigma^4 I$.

Since $\lim_{S, N \rightarrow \infty, \text{seq}} N^{-1} E(\varepsilon_{i,k}^{2l} u_{i,s}^4 (\Delta u_{i,t})^{2q}) < \infty \forall k, s, t$ and $\forall l, q \geq 0$ by lemma 10, we have a fortiori $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-1} \text{Var}([y_{i,1} \Delta y_{i,t}]^2) < \infty$ and $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-1} \text{Var}([y_{i,1} \Delta v_{i,t}]^2) < \infty \forall t$ when $d \geq 0$. Moreover, when $d = 0$, we have $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} \text{Var}([y_{i,1} \Delta y_{i,t}]^2) < \infty$ and $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} \text{Var}([y_{i,1} \Delta v_{i,t}]^2) < \infty \forall t$. It follows that $\text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-(1+\underline{d})} \sum_{i=1}^N Z_i^{I'} (\Delta y_i - \Delta y_{i,-1}) (\Delta y_i - \Delta y_{i,-1})' Z_i^I = \text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-(1+\underline{d})} \sum_{i=1}^N Z_i^{I'} \Delta v_i \Delta v_i' Z_i^I = (\sigma^2 \bar{q}(\lambda, \hat{c}) / (2\lambda) + 1\{d=0\} \sigma_{-S}^2) \sigma^2 H$, and $\text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-(1+\frac{1}{2}\underline{d})} \sum_{i=1}^N Z_i^{I'} (\Delta y_i - \Delta y_{i,-1}) \Delta y_i' Z_i^{II} = \text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-(1+\frac{1}{2}\underline{d})} \times \sum_{i=1}^N (Z_i^{I'} (\Delta v_i) v_i' Z_i^{II}) = 0$, where $\underline{d} = \min(d, 1/2)$, $1\{d=0\} = 1$ if $d = 0$ and $1\{d=0\} = 0$ if $d \neq 0$, $\sigma_{-S}^2 = E(y_{i,-S}^2)$ and $\bar{q}(\lambda, \hat{c}) / (2\lambda) \equiv \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-d} (1 - \rho^{2(1+S)}) / (1 - \rho^2)$ (cf $\hat{q}(\lambda, \hat{c}) / (2\lambda)$ in the proof of theorem 5). Note that $\bar{q}(\lambda, \hat{c}) / (2\lambda) = S+1$ if $d = 0$; $\bar{q}(\lambda, \hat{c}) / (2\lambda) = \hat{c}$ if $0 < d < 1/2$; $\bar{q}(\lambda, \hat{c}) / (2\lambda) = q(\lambda, \hat{c}) / (2\lambda)$ if $d = 1/2$; and $\bar{q}(\lambda, \hat{c}) / (2\lambda) = 1 / (2\lambda)$ if $d > 1/2$.

Next we consider second moments. When $0 < d < 1/2$, we have $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} E((1 - \rho) u_{i,1}^2) = \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} \sigma^2 (1 - \rho^{2(1+S)}) / 2 = 0$ by part (b) of lemma 11. When $d = 0$, we obtain $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} E((1 - \rho) u_{i,1}^2) = 0$ as well. Finally, when $d \geq 1/2$, we have $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} E((1 - \rho) u_{i,1}^2) = \lim_{N \rightarrow \infty, S/N^{1/2} \rightarrow c, S/N^d \rightarrow \hat{c}} \sigma^2 (1 - \rho^{2(1+S)}) / 2 = \sigma^2 q(\lambda, c) / 2$.

Recall that if $c = 0$, then $q(\lambda, c) = 0$. Then it is easily verified that for $d \geq 0$ $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} E(u_{i,t} \Delta u_{i,t}) = \sigma^2 + \lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} E(u_{i,t-1} \Delta u_{i,t}) = \sigma^2 + \lim_{N \rightarrow \infty, S/N^{1/2} \rightarrow c, S/N^d \rightarrow \hat{c}} E((\rho-1) u_{i,1}^2) = \sigma^2 - \sigma^2 q(\lambda, c) / 2$. Moreover, since $\lim_{S, N \rightarrow \infty, \text{seq}} N^{-1} \times E(u_{i,t}^2 (\Delta u_{i,t})^2) = 0$ by lemma 10, we have a fortiori $\lim_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-1} E(u_{i,t}^2 (\Delta u_{i,t})^2) = 0$ when $d \geq 0$. It follows that $N^{-1} \sum_{i=1}^N Z_i^{II'} y_{i,-1} \xrightarrow{q.m.} \sigma^2 (1 - q(\lambda, c) / 2) \iota$ and hence $\text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-1} \sum_{i=1}^N Z_i^{II'} y_{i,-1} \equiv \vec{X}_8 = \sigma^2 (1 - q(\lambda, c) / 2) \iota$. Similarly, we obtain that $\text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} N^{-(1+\frac{1}{2}\underline{d})} \sum_{i=1}^N Z_i^{I'} \Delta y_{i,-1} = 0$ when $d \geq 0$.

Finally, note that $N^{-\frac{1}{2}(1+\underline{d})} \sum_{i=1}^N Z_i^{I'} \Delta v_i \xrightarrow{d} \vec{X}_{51} \sim N(0, (\sigma^2 \bar{q}(\lambda, \hat{c}) / (2\lambda) + 1\{d=0\} \sigma_{-S}^2) \times \sigma^2 H)$ and $N^{-1/2} \sum_{i=1}^N Z_i^{II'} v_i \xrightarrow{d} \vec{X}_7 \sim N(0, \sigma^4 I)$ as $N, S \rightarrow \infty$ with $S/N^{1/2} \rightarrow c \geq 0$ and $S/N^d \rightarrow \hat{c} > 0$, where $d \geq 0$.

The above results imply that $\text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} [SE(\hat{\rho}_{Arbov2,R})]^{-2} / N = \vec{X}_8' \vec{X}_8 / \sigma^4 = (1 - q(\lambda, c) / 2)^2 m$ and $\text{plim}_{N \rightarrow \infty, S/N^d \rightarrow \hat{c}} [SE(\hat{\rho}_{SYS2,R})]^{-2} / N = \vec{X}_8' \vec{X}_8 / \sigma^4$. Moreover, as $N, S \rightarrow \infty$ with $S/N^{1/2} \rightarrow c \geq 0$ and $S/N^d \rightarrow \hat{c} > 0$, where $d \geq 0$, $(\hat{\rho}_{Arbov2,R} - \rho) / SE(\hat{\rho}_{Arbov2,R}) \xrightarrow{d} \sigma^{-2} (\vec{X}_8' \vec{X}_8)^{-1/2} \vec{X}_8' \vec{X}_7$ and $(\hat{\rho}_{SYS2,R} - \rho) / SE(\hat{\rho}_{SYS2,R}) \xrightarrow{d} \sigma^{-2} (\vec{X}_8' \vec{X}_8)^{-1/2} \vec{X}_8' \vec{X}_7$.

Note that $\vec{X}_7 \perp \vec{X}_8$ Therefore $\sigma^{-2} (\vec{X}_8' \vec{X}_8)^{-1/2} \vec{X}_8' \vec{X}_7 | \vec{X}_8 \sim N(0, 1)$. We conclude that

$(\hat{\rho}_{Arbov2,R}-1)/SE(\hat{\rho}_{Arbov2,R}) \xrightarrow{d} N(-\lambda(1-q(\lambda,c)/2)\sqrt{m}, 1)$ and $(\hat{\rho}_{SYS2,R}-1)/SE(\hat{\rho}_{SYS2,R}) \xrightarrow{d} N(-\lambda(1-q(\lambda,c)/2)\sqrt{m}, 1)$. \square

References

- [1] Ahn, S.C., and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, *Journal of Econometrics* 68, 5-28.
- [2] 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.
- [3] 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.
- [4] Arellano, M., and O. Bover, 1995, Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics* 68, 29-51.
- [5] Blundell, R.W., and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* 87, 115-143.
- [6] Bond, S.R., and F. Windmeijer, 2002, Finite sample inference for GMM estimators in linear panel data models, CEMMAP working paper CWP04/02, IFS, London.
- [7] Bond, S.R., C. Nauges, and F. Windmeijer, 2005, Unit roots: identification and testing in micro panels, CEMMAP working paper CWP07/05, IFS, London.
- [8] 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.
- [9] Fieller, E.C., 1932, The distribution of an index in a normal bivariate population, *Biometrika* 24, 428-440.
- [10] Hahn, J., and J. Hausman, 2002, Notes on bias in estimators for simultaneous equation models, *Economics Letters* 75, 237-241.
- [11] Hahn, J., J. Hausman, and G. Kuersteiner, 2001, Bias corrected instrumental variables

- estimation for dynamic panel models with fixed effects, unpublished manuscript, MIT, Cambridge, Massachusetts.
- [12] Han, C., and P.C.B. Phillips, 2006, GMM with many moment conditions, *Econometrica* 74, 147-192.
- [13] 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.
- [14] Hsiao, C., M.H. Pesaran, and A.K. Tahmiscioglu, 2002, Maximum likelihood estimation of fixed effects dynamic panel models covering short time periods, *Journal of Econometrics* 109, 107-150.
- [15] Kruiniger, H., 2000, GMM estimation of dynamic panel data models with persistent data, Working paper 428, Queen Mary, University of London, revised in December 2003.
- [16] Kruiniger, H., 2001, On the estimation of panel regression models with fixed effects, Working paper 450, Queen Mary, University of London, revised in 2003.
- [17] Kruiniger, H., 2004, Maximum likelihood estimation and inference methods for the stationary panel AR(1)/UR model, appeared earlier as working paper 429, Queen Mary, University of London.
- [18] Kruiniger, H., 2005, Improved GMM estimation and Wald inference in dynamic panel data models under alternative asymptotics, unpublished manuscript, Queen Mary, University of London.
- [19] Madsen, E., 2003, Using GMM when testing for a unit root in panels where the time-dimension is fixed, CAM working paper 2003-11, University of Copenhagen.
- [20] Moon, H.R., and P.C.B. Phillips, 2000, Estimation of autoregressive roots near unity using panel data, *Econometric Theory* 16, 927-997.
- [21] Phillips, P.C.B., and H.R. Moon, 1999, Linear regression limit theory for nonstationary panel data, *Econometrica* 67, 1057-1111.

- [22] Staiger, D., and J.H. Stock, 1997, Instrumental variables regression with weak instruments, *Econometrica* 65, 557-586.
- [23] Windmeijer, F., 2005, A finite sample correction for the variance of linear efficient two-step GMM estimators, *Journal of Econometrics*, 25-51.

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

**Copyright © 2006 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
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm**