An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis*

M. Hashem Pesaran
Trinity College, Cambridge, England

Yongcheol Shin
Department of Applied Economics, University of Cambridge, England

First Version: February, 1995, Revised: January, 1997

Abstract

This paper examines the use of autoregressive distributed lag (ARDL) models for the analysis of long-run relations when the underlying variables are I(1). It shows that after appropriate augmentation of the order of the ARDL model, the OLS estimators of the short-run parameters are $\sqrt{T}$-consistent with the asymptotically singular covariance matrix, and the ARDL-based estimators of the long-run coefficients are super-consistent, and valid inferences on the long-run parameters can be made using standard normal asymptotic theory. The paper also examines the relationship between the ARDL procedure and the fully modified OLS approach of Phillips and Hansen to estimation of cointegrating relations, and compares the small sample performance of these two approaches via Monte Carlo experiments. These results provide strong evidence in favour of a rehabilitation of the traditional ARDL approach to time series econometric modelling. The ARDL approach has the additional advantage of yielding consistent estimates of the long-run coefficients that are asymptotically normal irrespective of whether the underlying regressors are I(1) or I(0).

JEL Classifications: C12, C13, C15, C22.
Key Words: Autoregressive distributed lag model, Cointegration, I(1) and I(0) regressors, Model selection, Monte Carlo simulation.

---

*This is a revised version of a paper presented at the Symposium at the Centennial of Ragnar Frisch, The Norwegian Academy of Science and Letters, Oslo, March 3-5, 1995. We are grateful to Peter Boswijk, Clive Granger, Alberto Holly, Kyung So Im, Brendan McCabe, Steve Satchell, Richard Smith, Ron Smith and an anonymous referee for helpful comments. Partial financial support from the ESRC (Grant No. R000233608) and the Isaac Newton Trust of Trinity College, Cambridge is gratefully acknowledged.

1. INTRODUCTION

Econometric analysis of long-run relations has been the focus of much theoretical and empirical research in economics. In the case where the variables in the long-run relation of interest are trend stationary, the general practice has been to de-trend the series and to model the de-trended series as stationary distributed lag or autoregressive distributed lag (ARDL) models. Estimation and inference concerning the long-run properties of the model are then carried out using standard asymptotic normal theory. (For a comprehensive review of this literature see Hendry, Pagan and Sargan (1984) and Wickens and Breusch (1988)). The analysis becomes more complicated when the variables are difference-stationary, or integrated of order 1 (I(1) for short). The recent literature on cointegration is concerned with the analysis of the long run relations between I(1) variables, and its basic premise is, at least implicitly, that in the presence of I(1) variables the traditional ARDL approach is no longer applicable. Consequently, a large number of alternative estimation and hypothesis testing procedures have been specifically developed for the analysis of I(1) variables. (See the pioneering work of Engle and Granger (1987), Johansen (1991), Phillips (1991), Phillips and Hansen (1990) and Phillips and Loretan (1991).)

In this paper we re-examine the use of the traditional ARDL approach for the analysis of long run relations when the underlying variables are I(1). We consider the following general ARDL(p, q) model:

$$y_t = \alpha_0 + \alpha_1 t + \sum_{i=1}^{p} \phi_i y_{t-i} + \boldsymbol{\beta}'\mathbf{x}_t + \sum_{i=0}^{q-1} \boldsymbol{\beta}_i^{*\prime} \Delta\mathbf{x}_{t-i} + u_t, \tag{1.1}$$

$$\Delta\mathbf{x}_t = \mathbf{P}_1 \Delta\mathbf{x}_{t-1} + \mathbf{P}_2 \Delta\mathbf{x}_{t-2} + \cdots + \mathbf{P}_s \Delta\mathbf{x}_{t-s} + \boldsymbol{\varepsilon}_t, \tag{1.2}$$

where $\mathbf{x}_t$ is the k-dimensional I(1) variables that are not cointegrated among themselves, u_t and $\boldsymbol{\varepsilon}_t$ are serially uncorrelated disturbances with zero means and constant variance-covariances, and $\mathbf{P}_i$ are $k \times k$ coefficient matrices such that the vector autoregressive process in $\Delta\mathbf{x}_t$ is stable. We also assume that the roots of $1 - \sum_{i=1}^{p} \phi_i z^i = 0$ all fall outside the unit circle and there exists a stable unique long-run relationship between y_t and $\mathbf{x}_t$.

We consider the problem of consistent estimation of the parameters of the ARDL model both when u_t and $\boldsymbol{\varepsilon}_t$ are uncorrelated, and when they are correlated. In the former case we will show that the OLS estimators of the short-run parameters, α_0, α_1, $\boldsymbol{\beta}$, $\boldsymbol{\beta}_1^*, ..., \boldsymbol{\beta}_{q-1}^*$ and $\boldsymbol{\phi} = (\phi_1, ..., \phi_p)$ are $\sqrt{T}$- consistent, and the covariance matrix of these estimators has a well-defined limit which is asymptotically singular such that the estimators of α_1 and $\boldsymbol{\beta}$ are asymptotically perfectly collinear with the estimator of $\boldsymbol{\phi}$. These results have the interesting implication that the OLS estimators of the long-run coefficients, defined by the ratios $\delta = \alpha_1/\phi(1)$ and $\boldsymbol{\theta} = \boldsymbol{\beta}/\phi(1)$, where $\phi(1) = 1 - \sum_{i=1}^{p} \phi_i$, converge to their true values faster than the estimators of the short run parameters α_1 and $\boldsymbol{\beta}$. The ARDL-based estimators of δ and $\boldsymbol{\theta}$ are $T^{\frac{3}{2}}$-consistent and T-consistent, respectively. These results are not surprising and are familiar from the cointegration literature. But more importantly, we will show that despite the singularity of the covariance structure of the OLS estimators of the short-run parameters, valid inferences on δ and $\boldsymbol{\theta}$, as well as on individual short run parameters, can be made using standard normal asymptotic theory. Therefore, the traditional ARDL approach justified in the case of trend-stationary regressors, is in fact equally valid even if the regressors are first-difference stationary.

In the case where u_t and $\boldsymbol{\varepsilon}_t$ are correlated the ARDL specification needs to be augmented with an adequate number of lagged changes in the regressors before estimation and inference are carried out. The degree of augmentation required depends on whether $q > s+1$ or not. Denoting the contemporaneous correlation between u_t and $\boldsymbol{\varepsilon}_t$ by the $k \times 1$ vector $\mathbf{d}$, the augmented version of (1.1) can be written as

$$y_t = \alpha_0 + \alpha_1 t + \sum_{i=1}^{p} \phi_i y_{t-i} + \boldsymbol{\beta}'\mathbf{x}_t + \sum_{i=0}^{m-1} \boldsymbol{\pi}_i' \Delta \mathbf{x}_{t-i} + \eta_t, \qquad (1.3)$$

where $m = max(q, s+1)$, $\boldsymbol{\pi}_i = \boldsymbol{\beta}_i^* - \mathbf{P}_i'\mathbf{d}$, $i = 0, 1, 2, ..., m-1$, $\mathbf{P}_0 = \mathbf{I}_k$, where $\mathbf{I}_k$ is a $k \times k$ identity matrix, $\boldsymbol{\beta}_i^* = 0$ for $i \geq q$, and $\mathbf{P}_i = 0$ for $i \geq s$. In this augmented specification η_t and $\boldsymbol{\varepsilon}_t$ are uncorrelated and the results stated above will be directly applicable to the OLS estimators of the short-run and long-run parameters of (1.3). Once again traditional methods of estimation and inference, originally developed for trend-stationary variables, are applicable to first-difference stationary variables. The estimation of the short-run effects still requires an explicit modelling of the contemporaneous dependence between u_t and $\boldsymbol{\varepsilon}_t$. In practice, an appropriate choice of the order of the ARDL model is crucial for valid inference. But once this is done, estimation of the long-run parameters and computation of valid standard errors for the resultant estimators can be carried out either by the OLS method, using the so-called "delta" method (Δ-method) to compute the standard errors, or by the Bewely's (1979) regression approach. These two procedures yield identical results and a choice between them is only a matter of computational convenience.

The use of the ARDL estimation procedure is directly comparable to the semi-parametric, fully-modified OLS approach of Phillips and Hansen (1990) to estimation of cointegrating relations. In the static formulation of the cointegrating regression,

$$y_t = \mu + \delta t + \boldsymbol{\theta}'\mathbf{x}_t + v_t, \qquad (1.4)$$

where $\Delta \mathbf{x}_t = \mathbf{e}_t$, and $\boldsymbol{\xi}_t = (v_t, \mathbf{e}'_t)'$ follows a general linear stationary process, the OLS estimators of δ and $\boldsymbol{\theta}$ are $T^{\frac{3}{2}}$- and T-consistent, but in general the asymptotic distribution of the OLS estimator of $\boldsymbol{\theta}$ involves the unit-root distribution as well as the second-order bias in the presence of the contemporaneous correlations that may exist between v_t and $\mathbf{e}_t$. Therefore, the finite sample performance of the OLS estimator is poor and in addition, due to the nuisance parameter dependencies, inference on $\boldsymbol{\theta}$ using the usual t-tests in the OLS regression of (1.4) is invalid. To overcome these problems Phillips and Hansen (1990) have suggested the fully-modified OLS estimation procedure that asymptotically takes account of these correlations in a semi-parametric manner, in the sense that the fully-modified estimators have the Gaussian mixture normal distribution asymptotically, and inferences on the long run parameters using the t-test based on the limiting distribution of the fully-modified estimator is valid.

The ARDL-based approach to estimation and inference, and the fully-modified OLS procedure are both asymptotically valid when the regressors are I(1), and a choice between them has to be made on the basis of their small sample properties and computational convenience. To examine the small sample performance of the two estimators we have carried out a number of Monte Carlo experiments. Since in practice the "true" orders of the ARDL(p, m) model are rarely known *a priori*, in the Monte Carlo experiments we also consider a two-step strategy whereby p and m are first selected (estimated) using either the Akaike Information Criterion (AIC), or the Schwarz Bayesian Criterion (SC), and then the long-run coefficients and their standard errors are estimated using the ARDL model selected in the first step. We refer to these estimators as ARDL-AIC and ARDL-SC. The main findings from these experiments are as follows:

(i) The ARDL-AIC and the ARDL-SC estimators have very similar small-sample performances, with the ARDL-SC performing slightly better in the majority of the experiments. This may reflect the fact that the Schwartz criterion is a consistent model selection criterion while Akaike is not.

(ii) The ARDL test statistics that are computed using the Δ-method (or equivalently by means of the so-called Bewley's regression), generally perform much better in small samples than the test statistics computed using the asymptotic formula that explicitly takes account of the fact that the regressors are I(1).

(iii) The ARDL-SC procedure when combined with the Δ-method of computing the standard errors of the long-run parameters generally dominates the Phillips-Hansen estimator in small samples. This is in particular true of the size-power performance of the tests on the long-run parameter.

(iv) The Monte Carlo results point strongly in favor of the two-step estimation procedure, and this strategy seems to work even when the model under consideration has endogenous regressors, irrespective of whether the regressors are I(1) or I(0).[1]

The plan of the paper is as follows: Section 2 examines the asymptotic properties of the OLS estimators in the context of a simple autoregressive model with a linear deterministic trend and the k-dimensional strictly exogenous I(1) regressors. Section 3 considers a more general ARDL model, allowing for residual serial correlations and possible endogeneity of the I(1) regressors, and develops the resultant asymptotic theory. In Section 4 the ARDL-based approach is compared to the cointegration-based approach of Phillips and Hansen (1990). Section 5 reports and discusses the results of Monte Carlo experiments. Some concluding remarks are presented in Section 6. Mathematical proofs are provided in an Appendix.

2. The Lagged Dependent Variable Model with the Deterministic Trend and Exogenous I(1) Regressors

Initially we consider the simple ARDL(1,0) model containing I(1) regressors and a linear deterministic trend,

$$\phi(L)y_t = \alpha_0 + \alpha_1 t + \boldsymbol{\beta}'\mathbf{x}_t + u_t,\ t = 1, ..., T, \tag{2.1}$$

where y_t is a scalar, $\phi(L) = 1 - \phi L$, with L being the one period lag operator, $\mathbf{x}_t$ is a $k \times 1$ vector of regressors assumed to be integrated of order 1:[2]

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{e}_t, \tag{2.2}$$

and $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters. Suppose that the following assumptions hold:

(A1) The scalar disturbance term, u_t, in (2.1) is $iid(0, \sigma_u^2)$,

[1]The case where the regressors are I(1) and cointegrated among themselves presents additional identification problems and is best analyzed in the context of a system of long-run structural equations. On this see Pesaran and Shin (1995).

[2]Specifications (2.1) and (2.2) can easily be adapted to allow for inclusion of a drift term in the $\mathbf{x}_t$ process. Consider, for example, the process $\Delta\mathbf{x}_t = \boldsymbol{\mu}_x + \mathbf{e}_t$, and note that it can also be written as $\mathbf{x}_t = \boldsymbol{\mu}_x t + \tilde{\mathbf{x}}_t$, where $\Delta\tilde{\mathbf{x}}_t = \mathbf{e}_t$. Therefore, substituting $\mathbf{x}_t$ in (2.1) we have

$$\phi(L)y_t = \alpha_0 + (\alpha_1 + \boldsymbol{\beta}'\boldsymbol{\mu}_x)t + \boldsymbol{\beta}'\tilde{\mathbf{x}}_t + u_t,$$

where $\tilde{\mathbf{x}}_t$ follows an I(1) process without a drift.

(A2) The k-dimensional vector, $\mathbf{e}_t$, in (2.2) has a general linear multivariate stationary process,

(A3) u_t and $\mathbf{e}_t$ are uncorrelated for all leads and lags such that $\mathbf{x}_t$ is strictly exogenous with respect to u_t,

(A4) The I(1) regressors, $\mathbf{x}_t$, are not cointegrated among themselves, and

(A5) $|\phi| < 1$, so that the model is dynamically stable, and a long-run relationship between y_t and $\mathbf{x}_t$ exists.[3]

We shall distinguish between two types of parameters, the parameters capturing the short-run dynamics ($\alpha_0, \alpha_1, \boldsymbol{\beta}$ and ϕ), and the long run parameters on the trended regressors, t and $\mathbf{x}_t$, defined by

$$\delta = \frac{\alpha_1}{1-\phi}, \ \boldsymbol{\theta} = \frac{\boldsymbol{\beta}}{1-\phi}. \tag{2.3}$$

Applying the decomposition $1 - \phi L = (1 - \phi) + \phi(1 - L)$ to (2.1), y_t can be expressed as

$$y_t = \mu + \delta t + \boldsymbol{\theta}'\mathbf{x}_t + v_t, \tag{2.4}$$

where

$$\mu = \frac{\alpha_0}{1-\phi} - \left(\frac{\phi}{1-\phi}\right)\delta,$$

and

$$v_t = \sum_{i=0}^{\infty} \phi^i u_{t-i} - \phi \sum_{i=0}^{\infty} \phi^i \boldsymbol{\theta}'\mathbf{e}_{t-i}.$$

From (2.1) and (2.4) it is clear that y_t and $\mathbf{x}_t$ are individually I(1), but must be cointegrated for (2.1) to be meaningful.[4] Similarly, we obtain

$$y_{t-1} = \mu_1 + \delta t + \boldsymbol{\theta}'\mathbf{x}_t + \kappa_t, \tag{2.5}$$

where $\mu_1 = \mu - \delta$, $\kappa_t = v_{t-1} - \boldsymbol{\theta}'\mathbf{e}_t$, and κ_t is an I(0) process with variance σ_κ^2.

Our main aim is to derive the asymptotic properties of the OLS estimators of the short-run as well as the long-run parameters in the context of the ARDL(1,0)

[3] Tests of the existence of long-run relationships between y_t and $\mathbf{x}_t$, when it is not known *a priori* whether $\mathbf{x}_t$ are I(0) or I(1), are discussed in Pesaran, Shin and Smith (1996).

[4] A relationship between I(1) variables is said to be "stochastically cointegrated" if it is trend stationary, while "deterministic cointegration" refers to the case where the cointegrating relation is level stationary. For a discussion of these two types of cointegrating relations see Park (1992).

model, (2.1). For expositional convenience, we transform (2.1) to the partitioned regression model in the matrix form as,

$$\mathbf{y}_T = \mathbf{Z}_T\mathbf{b} + \mathbf{y}_{T-1}\phi + \mathbf{u}_T, \tag{2.6}$$

where $\mathbf{y}_T = (y_1, ..., y_T)'$, $\mathbf{y}_{T-1} = (y_0, ..., y_{T-1})'$, $\boldsymbol{\tau}_T = (1, ..., 1)'$, $\mathbf{t}_T = (1, ..., T)'$, $\mathbf{X}_T = (\mathbf{x}_1, ..., \mathbf{x}_T)'$, $\mathbf{Z}_T = (\boldsymbol{\tau}_T, \mathbf{t}_T, \mathbf{X}_T)$, $\mathbf{u}_T = (u_1, ..., u_T)'$, and $\mathbf{b} = (\alpha_0, \alpha_1, \boldsymbol{\beta}')'$. Since our main interest is in the long-run coefficients on trended regressors, t and $\mathbf{x}_t$, we also partition

$$\mathbf{Z}_T = (\boldsymbol{\tau}_T, \mathbf{S}_T),\ \mathbf{S}_T = (\mathbf{t}_T, \mathbf{X}_T),\ \mathbf{b} = \begin{pmatrix} \alpha_0 \\ \mathbf{c} \end{pmatrix},\ \mathbf{c} = \begin{pmatrix} \alpha_1 \\ \boldsymbol{\beta} \end{pmatrix},$$

where the dimensions of $\mathbf{Z}_T$, $\mathbf{S}_T$, $\mathbf{b}$ and $\mathbf{c}$ are $T \times (k+2)$, $T \times (k+1)$, $(k+2) \times 1$ and $(k+1) \times 1$, respectively.

Theorem 2.1. *Under the assumptions (A1) - A(5), the OLS estimators of ϕ and $\mathbf{c} = (\alpha_1, \boldsymbol{\beta}')'$ in (2.6), denoted by $\hat{\phi}_T$ and $\hat{\mathbf{c}}_T$, respectively, are $\sqrt{T}$-consistent, and have the following asymptotic distributions:*

$$\sqrt{T}(\hat{\phi}_T - \phi) \overset{a}{\sim} N\left\{0, \frac{\sigma_u^2}{\sigma_\kappa^2}\right\}, \tag{2.7}$$

$$\sqrt{T}(\hat{\mathbf{c}}_T - \mathbf{c}) \overset{a}{\sim} N\left\{\mathbf{0}, \frac{\sigma_u^2}{\sigma_\kappa^2}\boldsymbol{\lambda}\boldsymbol{\lambda}'\right\}, \tag{2.8}$$

where $\boldsymbol{\lambda} = (\delta, \boldsymbol{\theta}')'$ is a $(k+1) \times 1$ vector of the long run parameters on trended regressors, t and $\mathbf{x}_t$, and $rank(\boldsymbol{\lambda}\boldsymbol{\lambda}') = 1$. In addition, the OLS estimator of α_0 in (2.6), denoted by $\hat{\alpha}_{0T}$, is also $\sqrt{T}$-consistent, but has the mixture normal distribution. Defining $\mathbf{h} = (\mathbf{b}', \phi)'$ and $\mathbf{P}_{Z_T} = (\mathbf{Z}_T, \mathbf{y}_{T-1})$, and denoting the OLS estimator of $\mathbf{h}$ by $\hat{\mathbf{h}}_T$, the covariance matrix of $\hat{\mathbf{h}}_T$ can be consistently estimated by

$$\hat{V}(\hat{\mathbf{h}}_T) = \hat{\sigma}_{uT}^2(\mathbf{P}_{Z_T}'\mathbf{P}_{Z_T})^{-1},$$

where $\hat{\sigma}_{uT}^2 = T^{-1}(\mathbf{y}_T - \mathbf{P}_{Z_T}\hat{\mathbf{h}}_T)'(\mathbf{y}_T - \mathbf{P}_{Z_T}\hat{\mathbf{h}}_T)$, and $\hat{V}(\hat{\mathbf{h}}_T)$ is asymptotically singular with rank equal to 2.

Theorem 2.1 shows that despite the presence of stochastic and deterministic trends in the ARDL model, the OLS estimators of the short-run parameters are $\sqrt{T}$-consistent.[5] The second and more important finding is that the OLS estimators

[5]Similar results can also be obtained in the case of regressors with higher order trend terms such as t^2, t^3, ..., or I(2), I(3), ..., variables.

of the coefficients on the trended regressors, α_1 and $\boldsymbol{\beta}$, in (2.1) are asymptotically perfectly collinear with the OLS estimator of the coefficient on the lagged dependent variable, ϕ; namely,

$$\sqrt{T}\left\{(\hat{\mathbf{c}}_T - \mathbf{c}) + \boldsymbol{\lambda}(\hat{\phi}_T - \phi)\right\} = o_p(1). \tag{2.9}$$

One interesting implication of this result is that the t-statistics for testing the significance of individual impact coefficients on the I(1) regressors are asymptotically equivalent, namely $t_{\hat{\beta}_i} - t_{\hat{\beta}_j} = o_p(1)$ for $i \neq j$, and $t_{\hat{\beta}_i} - t_{\hat{\alpha}_1} = o_p(1)$.[6] Furthermore, $t_{\hat{\beta}_i} - t_{(1-\hat{\phi})} = o_p(1)$. Relation (2.9) in conjunction with

$$\hat{\boldsymbol{\lambda}}_T - \boldsymbol{\lambda} = \frac{(\hat{\mathbf{c}}_T - \mathbf{c}) + \boldsymbol{\lambda}(\hat{\phi}_T - \phi)}{(1 - \hat{\phi}_T)}, \tag{2.10}$$

also yields an important result familiar from the cointegration literature, which we set out in the following theorem:

Theorem 2.2. *Under assumptions (A1) - (A5), the ARDL-based estimators of the long-run parameters, given by* $\hat{\delta}_T = \hat{\alpha}_{1T}/(1 - \hat{\phi}_T)$, *and* $\hat{\boldsymbol{\theta}}_T = \hat{\boldsymbol{\beta}}_T/(1 - \hat{\phi}_T)$, *converge to their true values* δ *and* $\boldsymbol{\theta}$, *respectively, at the rates,* $T^{\frac{3}{2}}$ *and* T. *Also asymptotically,* $T^{\frac{3}{2}}(\hat{\delta}_T - \delta)$ *and* $T(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta})$ *have the (mixture) normal distributions, and therefore,*

$$\mathbf{Q}_{\tilde{S}_T}^{\frac{1}{2}} \mathbf{D}_{S_T}^{-1}(\hat{\boldsymbol{\lambda}}_T - \boldsymbol{\lambda}) \overset{a}{\sim} N\left\{\mathbf{0}, \frac{\sigma_u^2}{(1-\phi)^2}\mathbf{I}_{k+1}\right\}, \tag{2.11}$$

where $\hat{\boldsymbol{\lambda}}_T = (\hat{\delta}_T, \hat{\boldsymbol{\theta}}_T')'$, $\mathbf{Q}_{\tilde{S}_T} = \mathbf{D}_{S_T}\mathbf{S}_T'\mathbf{H}_T\mathbf{S}_T\mathbf{D}_{S_T}$, $\mathbf{S}_T = (\mathbf{t}_T, \mathbf{X}_T)$, $\mathbf{H}_T = \mathbf{I}_T - \boldsymbol{\tau}_T(\boldsymbol{\tau}_T'\boldsymbol{\tau}_T)^{-1}\boldsymbol{\tau}_T'$, *and* $\mathbf{D}_{S_T} = Diag(T^{-\frac{3}{2}}, T^{-1}\mathbf{I}_k)$.

The finding that the estimator of $\boldsymbol{\theta}$ is T -consistent is known as the "super-consistency" property in the cointegration literature. Since the limiting distributions of $T^{\frac{3}{2}}(\hat{\delta}_T - \delta)$ and $T(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta})$ are (mixture) normal, optimal two-sided inferences concerning δ and $\boldsymbol{\theta}$ are possible. Notice also that the covariance matrix of the estimator of $\boldsymbol{\lambda}$ simply depends on the inverse of the (scaled) demeaned data matrix and the spectral density at zero frequency of $(1 - \phi L)^{-1}u_t$, namely $\sigma_u^2/(1-\phi)^2$. Once again, this finding is in line with the results already familiar from the cointegration literature. (See Section 4 for further discussions.)

[6] For large enough T we have $t_{\hat{\beta}_i} \approx (1-\phi)(\sigma_\kappa/\sigma_u)$. This explains the relatively low t-ratios often obtained for short-run coefficients in ARDL regressions with I(1) variables, especially when ϕ is close to unity.

Hypothesis testing on the general linear restrictions involving the $k+1$ dimensional long-run parameter vector, $\boldsymbol{\lambda}$, can be carried out in the usual manner. Consider the g linear restrictions on $\boldsymbol{\lambda}$,

$$\mathbf{R}\boldsymbol{\lambda} = \mathbf{r},$$

where $\mathbf{R}$ is a $g \times (k+1)$ matrix and $\mathbf{r}$ is a $g \times 1$ vector of known constants. These restrictions can be tested using the Wald statistic,

$$\begin{aligned} W &= (\mathbf{R}\hat{\boldsymbol{\lambda}}_T - \mathbf{r})' \left\{ \mathbf{R}Cov(\hat{\boldsymbol{\lambda}}_T)\mathbf{R}' \right\}^{-1} (\mathbf{R}\hat{\boldsymbol{\lambda}}_T - \mathbf{r}) \\ &= (\mathbf{R}\hat{\boldsymbol{\lambda}}_T - \mathbf{r})' \left\{ \frac{(1-\hat{\phi}_T)^2}{\hat{\sigma}^2_{uT}} (\mathbf{S}'_T\mathbf{H}_T\mathbf{S}_T) \right\} (\mathbf{R}\hat{\boldsymbol{\lambda}}_T - \mathbf{r}). \end{aligned} \tag{2.12}$$

Of special interest is the t-statistic on the individual coefficients given by

$$t_i = \frac{\hat{\lambda}_{iT} - \lambda_i}{\hat{s}_i}, \; i = 1, ..., k+1, \tag{2.13}$$

where the standard error of the i-th coefficient is consistently estimated by

$$\hat{s}_i = \sqrt{\frac{\hat{\sigma}^2_{uT}}{(1-\hat{\phi}_T)^2} (\mathbf{S}'_T\mathbf{H}_T\mathbf{S}_T)^{-1}_{ii}},$$

and $(\mathbf{S}'_T\mathbf{H}_T\mathbf{S}_T)^{-1}_{ii}$ denotes the i-th diagonal element of $(\mathbf{S}'_T\mathbf{H}_T\mathbf{S}_T)^{-1}$. By Theorem 2.2, the Wald statistic in (2.12) follows the asymptotic χ^2 distribution with g degrees of freedom, and t_i^2 in (2.13) is distributed asymptotically as a χ^2 variate with one degree of freedom.

It is worth noting that the results in Theorem 2.2 equally apply to the purely autoregressive model with deterministic trend,

$$y_t = \alpha_0 + \alpha_1 t + \phi y_{t-1} + u_t, \; t = 1, ..., T, \tag{2.14}$$

and to the ARDL(1,0) model without a deterministic trend,

$$y_t = \alpha_0 + \boldsymbol{\beta}'\mathbf{x}_t + \phi y_{t-1} + u_t, \; t = 1, ..., T. \tag{2.15}$$

For completeness the asymptotic results for these models are summarized in Theorems 2.3 and 2.4.

Theorem 2.3. *Under the assumptions (A1) and (A5), the OLS estimators of α_0, α_1 and ϕ in (2.14), denoted by $\hat{\alpha}_{0T}$, $\hat{\alpha}_{1T}$, and $\hat{\phi}_T$, are all $\sqrt{T}$-consistent, and asymptotically normally distributed. In addition, $\sqrt{T}(\hat{\alpha}_{1T} - \alpha_1)$ and $\sqrt{T}(\hat{\phi}_T - \phi)$*

are perfectly collinear asymptotically and the covariance matrix of $(\hat{\alpha}_{0T}, \hat{\alpha}_{1T}, \hat{\phi}_T)$ *is asymptotically singular with rank equal to 2. Furthermore, the estimator of the long run parameter* δ*, computed by* $\hat{\alpha}_{1T}/(1-\hat{\phi}_T)$*, has the following asymptotic distribution:*

$$T^{\frac{3}{2}}(\hat{\delta}_T - \delta) \overset{a}{\sim} N\left\{0, \frac{12\sigma_u^2}{(1-\phi)^2}\right\}. \tag{2.16}$$

Theorem 2.4. *Under assumptions (A1) - (A5), the OLS estimators of* α_0*,* β *and* ϕ *in (2.15), denoted by* $\hat{\alpha}_{0T}$*,* $\hat{\boldsymbol{\beta}}_T$*, and* $\hat{\phi}_T$ *are* $\sqrt{T}$*-consistent, and have the asymptotic (mixture) normal distributions. In addition,* $\sqrt{T}(\hat{\alpha}_{1T}-\alpha_1)$ *and* $\sqrt{T}(\hat{\phi}_T-\phi)$ *are perfectly collinear asymptotically and so the covariance matrix of* $(\hat{\alpha}_{0T}, \hat{\boldsymbol{\beta}}_T, \hat{\phi}_T)$ *is asymptotically singular with rank equal to 2. Furthermore, the estimator of the long run parameter* $\boldsymbol{\theta}$*, given by* $\hat{\boldsymbol{\theta}}_T = \hat{\boldsymbol{\beta}}_T/(1-\hat{\phi}_T)$*, has the mixture normal distribution asymptotically, and*

$$\mathbf{Q}_{\tilde{X}_T}^{\frac{1}{2}} T(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) \overset{a}{\sim} N\left\{\mathbf{0}, \frac{\sigma_u^2}{(1-\phi)^2}\mathbf{I}_k\right\}, \tag{2.17}$$

where $\mathbf{Q}_{\tilde{X}_T} = T^{-2}\mathbf{X}_T'\mathbf{H}_T\mathbf{X}_T$*.*

Before considering a more general specification of the ARDL model, we examine the relation between the standard errors of the estimator of the long-run parameter, $\boldsymbol{\theta}$, obtained from our asymptotic results and the standard errors obtained from the so called "delta" method (Δ-method for short). For ease of exposition we consider the simple model (2.15), and without loss of generality focus on the case where x_t is a scalar (i.e., $k=1$). From Theorem 2.4 we have

$$\mathbf{Q}_{\tilde{X}_T}^{\frac{1}{2}} T(\hat{\theta}_T - \theta) = \left[\sum_{t=1}^{T}(x_t-\bar{x})^2\right]^{\frac{1}{2}}(\hat{\theta}_T - \theta) \overset{a}{\sim} N\left\{0, \frac{\sigma_u^2}{(1-\phi)^2}\right\}, \tag{2.18}$$

where $\mathbf{Q}_{\tilde{X}_T} = T^{-2}\sum_{t=1}^{T}(x_t-\bar{x})^2$ and $\bar{x} = T^{-1}\sum_{t=1}^{T}x_t$.[7] Hence a consistent estimator of the variance of $\hat{\theta}_T$ is given by

$$\hat{V}(\hat{\theta}_T) = \frac{\hat{\sigma}_{uT}^2}{(1-\hat{\phi}_T)^2}\frac{1}{\sum_{t=1}^{T}(x_t-\bar{x})^2}. \tag{2.19}$$

[7]In the case where x_t is I(0) we have the same asymptotic result given by (2.18); that is, since $T^{-1}\mathbf{x}_T'\mathbf{H}_T\mathbf{x}_T = O_p(1)$ and $\sqrt{T}(\hat{\theta}_T-\theta) = O_p(1)$, hence

$$(T^{-1}\mathbf{x}_T'\mathbf{H}_T\mathbf{x}_T)^{\frac{1}{2}}\sqrt{T}(\hat{\theta}_T-\theta) = \left[\sum_{t=1}^{T}(x_t-\bar{x})^2\right]^{\frac{1}{2}}(\hat{\theta}_T-\theta) \overset{a}{\sim} N\left\{0, \frac{\sigma_u^2}{(1-\phi)^2}\right\}.$$

The computation of the variance of $\hat{\theta}_T$ by the Δ-method involves approximating

$$\hat{\theta}_T = g(\hat{\mathbf{\Psi}}_T) = \frac{\hat{\beta}_T}{1-\hat{\phi}_T},$$

by a linear function of $\hat{\mathbf{\Psi}}_T = (\hat{\beta}_T, \hat{\phi}_T)'$, and then approximating the variance of $\hat{\theta}_T$ by the variance of the resulting linear function. Denoting the estimator of the variance of $\hat{\theta}_T$ by $\hat{V}_\Delta(\hat{\theta}_T)$, we have

$$\begin{aligned}
\hat{V}_\Delta(\hat{\theta}_T) &= \left(\frac{\partial g(\hat{\mathbf{\Psi}}_T)}{\partial \hat{\mathbf{\Psi}}_T}\right)' \hat{V}(\hat{\mathbf{\Psi}}_T) \left(\frac{\partial g(\hat{\mathbf{\Psi}}_T)}{\partial \hat{\mathbf{\Psi}}_T}\right) \\
&= \left[\frac{1}{1-\hat{\phi}_T}, \frac{\hat{\beta}_T}{(1-\hat{\phi}_T)^2}\right] \hat{\sigma}^2_{uT}(\mathbf{R}_T'\mathbf{H}_T\mathbf{R}_T)^{-1} \begin{bmatrix} \frac{1}{1-\hat{\phi}_T} \\ \\ \frac{\hat{\beta}_T}{(1-\hat{\phi}_T)^2} \end{bmatrix},
\end{aligned}$$

where $\mathbf{R}_T = (\mathbf{x}_T, \mathbf{y}_{T-1})$. After some algebra $\hat{V}_\Delta(\hat{\theta}_T)$ can be expressed as

$$\hat{V}_\Delta(\hat{\theta}_T) = \frac{\hat{\sigma}^2_{uT}}{(1-\hat{\phi}_T)^2}\left[1, \hat{\theta}_T\right]\frac{1}{D_T}\begin{bmatrix} \sum(y_{t-1}-\bar{y})^2 & -\sum(y_{t-1}-\bar{y})(x_t-\bar{x}) \\ -\sum(y_{t-1}-\bar{y})(x_t-\bar{x}) & \sum(x_t-\bar{x})^2 \end{bmatrix}\begin{bmatrix} 1 \\ \hat{\theta}_T \end{bmatrix}, \tag{2.20}$$

where the bar over the variable denotes the sample mean, and

$$D_T = \left[\sum_{t=1}^T (x_t-\bar{x})^2\right]\left[\sum_{t=1}^T (y_{t-1}-\bar{y})^2\right] - \left[\sum_{t=1}^T (y_{t-1}-\bar{y})(x_t-\bar{x})\right]^2.$$

Using (2.5), recalling that $\delta = 0$ and defining $\tilde{y}_{t-1} = y_{t-1}-\bar{y}$, $\tilde{x}_t = x_t - \bar{x}$ and $\tilde{\kappa}_t = \kappa_t - \bar{\kappa}$, we also have

$$\tilde{y}_{t-1} = \theta\tilde{x}_t + \tilde{\kappa}_t, \tag{2.21}$$

where $\tilde{\kappa}_t$ follows a general linear stationary process. Substituting this result in (2.20), we obtain

$$\hat{V}_\Delta(\hat{\theta}_T) = \frac{\hat{\sigma}^2_{uT}}{(1-\hat{\phi}_T)^2}\frac{\sum_{t=1}^T \tilde{\kappa}_t^2 + (\hat{\theta}_T-\theta)^2\sum_{t=1}^T \tilde{x}_t^2 - 2(\hat{\theta}_T-\theta)\sum_{t=1}^T \tilde{x}_t\tilde{\kappa}_t}{(\sum_{t=1}^T \tilde{x}_t^2)(\sum_{t=1}^T \tilde{\kappa}_t^2) - (\sum_{t=1}^T \tilde{x}_t\tilde{\kappa}_t)^2}. \tag{2.22}$$

Since $\tilde{\kappa}_t$ is I(0) and $\tilde{x}_t$ is I(1), using the results familiar in the literature (see, for example, Phillips and Durlauf (1986)), we have

$$T^{-1}\sum_{t=1}^T \tilde{\kappa}_t^2 = O_p(1),\ T^{-2}\sum_{t=1}^T \tilde{x}_t^2 = O_p(1),\ T^{-1}\sum_{t=1}^T \tilde{x}_t\tilde{\kappa}_t = O_p(1).$$

Also from the result of Theorem 2.4 we know that $T(\hat{\theta}_T - \theta) = O_p(1)$. Hence, taking probability limits of the right hand side of (2.22) as $T \to \infty$, we have

$$\hat{V}_\Delta(\hat{\theta}_T) = \frac{\sigma_u^2}{(1-\phi)^2} \frac{1}{T^{-2}\sum_{t=1}^T (x_t - \bar{x})^2} + o_p(1).$$

Therefore, the standard error for the estimator of the long run parameter, θ, obtained using the Δ-method is asymptotically the same as that given by (2.19), which was derived assuming that x_t is I(1). *One important advantage of the variance estimator obtained by the Δ-method over the asymptotic formula (2.19) lies in the fact that it is asymptotically valid irrespective of whether x_t is I(1) or I(0), while the latter estimator is valid only if x_t is I(1).*

The two variance estimators clearly differ in finite samples. Notice that $(\sum_{t=1}^T \tilde{x}_t \tilde{\kappa}_t)^2$ is asymptotically negligible compared to other terms in (2.22), but it may not be negligible in finite samples, especially when $\tilde{x}_t$ and $\tilde{\kappa}_t$ are correlated. For a comparison of the small sample properties of the two variance estimators see the Monte Carlo results reported in Section 5.

3. General Autoregressive Distributed Lag Models with a Deterministic Trend and I(1) Regressors

So far we have derived the estimation and asymptotic results for the simple ARDL(1,0) model under the two strong assumptions (A1) and (A3). These assumptions, however, are too restrictive in the time series analysis, and so the estimation procedures developed in Section 2 are not expected to be robust to the violation of these assumptions, because the limiting distributions of the OLS estimators would then be inconsistent and/or depend on nuisance parameters.

We first relax the assumption (A1) and allow for the possibility of the error process in (2.1) to be serially correlated. To deal with this serial correlation we consider the ARDL(p, q) model,[8]

$$\phi(L)y_t = \alpha_0 + \alpha_1 t + \boldsymbol{\beta}'(L)\mathbf{x}_t + u_t, \tag{3.1}$$

where $\phi(L) = 1 - \sum_{j=1}^p \phi_j L^j$, and $\boldsymbol{\beta}(L) = \sum_{j=0}^q \boldsymbol{\beta}_j L^j$, and assume

(A1)$'$ The scalar disturbance, u_t, in the ARDL(p, q) model (3.1) is $iid(0, \sigma_u^2)$.

[8] For convenience we use the same notation u_t for the disturbance terms in (2.1) and (3.1). In practice the order of the lag polynomials operating on different elements of $\mathbf{x}_t$ could be different. But this does not affect the asymptotic theory presented below.

Using the decomposition $\boldsymbol{\beta}(L) = \boldsymbol{\beta}(1) + (1-L)\boldsymbol{\beta}^*(L)$, where $\boldsymbol{\beta}(1) = \sum_{j=0}^{q} \boldsymbol{\beta}_j$, $\boldsymbol{\beta}^*(L) = \sum_{j=0}^{q-1} \boldsymbol{\beta}_j^* L^j$ and $\boldsymbol{\beta}_j^* = -\sum_{i=j+1}^{q} \boldsymbol{\beta}_i$, (3.1) can be rewritten as

$$\phi(L)y_t = \alpha_0 + \alpha_1 t + \boldsymbol{\beta}'\mathbf{x}_t + \sum_{j=0}^{q-1} \boldsymbol{\beta}_j^{*\prime} \Delta\mathbf{x}_{t-j} + u_t, \tag{3.2}$$

where we have used $\boldsymbol{\beta} = \boldsymbol{\beta}(1)$. Similarly, applying the decomposition $\phi(L) = \phi(1) + (1-L)\phi^*(L)$ to (3.2), where $\phi(1) = 1 - \sum_{i=1}^{p} \phi_i$, $\phi^*(L) = \sum_{j=0}^{p-1} \phi_j^* L^j$ and $\phi_j^* = \sum_{i=j+1}^{p} \phi_i$, we have

$$\phi(1)y_t = \alpha_0 + \alpha_1 t + \boldsymbol{\beta}'\mathbf{x}_t + \sum_{j=0}^{q-1} \boldsymbol{\beta}_j^{*\prime} \Delta\mathbf{x}_{t-j} - \phi^*(L)\Delta y_t + u_t. \tag{3.3}$$

Also from (3.1), we obtain

$$\Delta y_t = [\phi(L)]^{-1} \{\alpha_1 + \boldsymbol{\beta}'(L)\Delta\mathbf{x}_t + \Delta u_t\}.$$

Substituting for Δy_t in (3.3) we have

$$y_t = \mu_0 + \delta t + \boldsymbol{\theta}'\mathbf{x}_t + \frac{\left\{\boldsymbol{\beta}^*(L) - \phi^*(L)\,[\phi(L)]^{-1}\,\boldsymbol{\beta}(L)\right\}'}{\phi(1)}\Delta\mathbf{x}_t + \frac{\left\{1 - (1-L)\phi^*(L)\,[\phi(L)]^{-1}\right\}}{\phi(1)}u_t, \tag{3.4}$$

where

$$\mu_0 = \frac{\alpha_0 - \phi^*(1)\delta}{\phi(1)}, \;\; \delta = \frac{\alpha_1}{\phi(1)}, \;\; \boldsymbol{\theta} = \boldsymbol{\theta}(1) = \frac{\boldsymbol{\beta}}{\phi(1)}.$$

Now it is easily seen that

$$\frac{(1-L)\boldsymbol{\beta}^*(L) - (1-L)\phi^*(L)\,[\phi(L)]^{-1}\,\boldsymbol{\beta}(L)}{\phi(1)} = \boldsymbol{\theta}(L) - \boldsymbol{\theta},$$

and

$$\frac{1 - (1-L)\phi^*(L)\,[\phi(L)]^{-1}}{\phi(1)} = \frac{1 - \{\phi(L) - \phi(1)\}\,[\phi(L)]^{-1}}{\phi(1)} = [\phi(L)]^{-1},$$

where $\boldsymbol{\theta}(L) = \boldsymbol{\beta}(L)/\phi(L)$. Using these results and the decomposition $\boldsymbol{\theta}(L) = \boldsymbol{\theta}(1) + (1-L)\boldsymbol{\theta}^*(L)$, where $\boldsymbol{\theta}^*(L) = \sum_{j=0}^{\infty} \boldsymbol{\theta}_j^* L^j$ and $\boldsymbol{\theta}_j^* = -\sum_{i=j+1}^{\infty} \boldsymbol{\theta}_i$ in (3.4) we obtain

$$y_t = \mu_0 + \delta t + \boldsymbol{\theta}'\mathbf{x}_t + \boldsymbol{\theta}^{*\prime}(L)\Delta\mathbf{x} + [\phi(L)]^{-1}\,u_t. \tag{3.5}$$

Matching the regressors on the right-hand-side of (3.2) with those in (3.5) we finally obtain

$$y_t = \mu_0 + \delta t + \boldsymbol{\theta}'\mathbf{x}_t + \sum_{j=0}^{q-1} \boldsymbol{\theta}_j^{*\prime} \Delta\mathbf{x}_{t-j} + \kappa_{0t}, \tag{3.6}$$

where $\kappa_{0t} = \sum_{j=q}^{\infty} \boldsymbol{\theta}_j^{*\prime} \mathbf{e}_{t-j} + [\phi(L)]^{-1} u_t$. Similarly,

$$y_{t-i} = \mu_i + \delta t + \boldsymbol{\theta}'\mathbf{x}_t + \sum_{j=0}^{q-1} \mathbf{g}_{ij}' \Delta \mathbf{x}_{t-j} + \kappa_{it},\ i = 1, ..., p, \tag{3.7}$$

where $\mu_i = \mu_0 - i\delta$, $i = 1, ..., p$,

$$\mathbf{g}_{ij} = \left\{ \begin{array}{ll} -\boldsymbol{\theta} & \text{if } i > j \\ \boldsymbol{\theta}_{j-1}^{*} & \text{if } i \leq j \end{array} \right\},\ 0 \leq j \leq q-1,\ i = 1, ..., p,$$

and

$$\kappa_{it} = \left\{ \begin{array}{cl} \sum_{j=q-i}^{\infty} \boldsymbol{\theta}_j^{*\prime} \mathbf{e}_{t-i-j} + [\phi(L)]^{-1} u_{t-i} & \text{for } i \leq q \\ -\boldsymbol{\theta}' \sum_{j=0}^{i-q-1} \mathbf{e}_{t-q-j} + \boldsymbol{\theta}^{*\prime}(L)\mathbf{e}_{t-i} + [\phi(L)]^{-1} u_t & \text{for } i > q \end{array} \right\}. \tag{3.8}$$

As in the previous section, we rewrite the ARDL(p, q) model (3.2) in matrix notations in the partitioned regression form,

$$\begin{aligned} \mathbf{y}_T &= \mathbf{G}_T \mathbf{f} + \mathbf{Y}_T \boldsymbol{\phi} + \mathbf{u}_T \qquad (3.9) \\ &= \alpha_0 \boldsymbol{\tau}_T + \mathbf{S}_T \mathbf{c} + \mathbf{W}_T \boldsymbol{\beta}^* + \mathbf{Y}_T \boldsymbol{\phi} + \mathbf{u}_T, \end{aligned}$$

where $\mathbf{y}_T = (y_1, ..., y_T)'$, $\mathbf{y}_{T,-i} = (y_{1-i}, ..., y_{T-i})'$, for $i = 1, ..., p$, $\mathbf{Y}_T = (\mathbf{y}_{T,-1}, ..., \mathbf{y}_{T,-p})$, $\Delta \mathbf{X}_{T,-j} = (\Delta \mathbf{x}_{1-j}, ..., \Delta \mathbf{x}_{T-j})$ for $j = 0, ..., q-1$, $\mathbf{W}_T = (\Delta \mathbf{x}_{T,0}, \Delta \mathbf{x}_{T,-1}, ..., \Delta \mathbf{x}_{T,-q+1})$, $\boldsymbol{\tau}_T = (1, ..., 1)'$, $\mathbf{t}_T = (1, ..., T)'$, $\mathbf{X}_T = (\mathbf{x}_1, ..., \mathbf{x}_T)'$, $\mathbf{G}_T = (\boldsymbol{\tau}_T, \mathbf{t}_T, \mathbf{X}_T, \mathbf{W}_T) = (\boldsymbol{\tau}_T, \mathbf{S}_T, \mathbf{W}_T)$, $\mathbf{u}_T = (u_1, ..., u_T)'$, $\mathbf{f} = (\alpha_0, \mathbf{c}', \boldsymbol{\beta}^{*\prime})'$, $\mathbf{c} = (\alpha_1, \boldsymbol{\beta}')'$, $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_0^{*\prime}, ..., \boldsymbol{\beta}_{q-1}^{*\prime})'$ and $\boldsymbol{\phi} = (\phi_1, ..., \phi_p)'$. Note that the dimensions of $\mathbf{Y}_T$, $\mathbf{G}_T$, $\boldsymbol{\phi}$ and $\mathbf{f}$ are $T \times p$, $T \times (k + kq + 2)$, $p \times 1$ and $(k + kq + 2) \times 1$, respectively.

Theorem 3.1. *Under assumptions (A1)′ and (A2) - (A5), the OLS estimators of* $\boldsymbol{\phi}$ *and* $\mathbf{c} = (\alpha_1, \boldsymbol{\beta}')'$ *in the ARDL(p, q) model (3.9) are* $\sqrt{T}$*-consistent and have the following asymptotic distributions:*

$$\sqrt{T}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}) \overset{a}{\sim} N\left\{\mathbf{0}, \sigma_u^2 \mathbf{Q}_K^{-1}\right\}, \tag{3.10}$$

where $\mathbf{Q}_K$ *is the* $p \times p$ *positive definite covariance matrix of* $(\kappa_{1t}, \kappa_{2t}, ..., \kappa_{pt})'$ *defined by (3.8), and*

$$\sqrt{T}(\hat{\mathbf{c}}_T - \mathbf{c}) \overset{a}{\sim} N\left\{\mathbf{0}, \sigma_u^2 \boldsymbol{\tau}_p' \mathbf{Q}_K^{-1} \boldsymbol{\tau}_p \boldsymbol{\lambda}\boldsymbol{\lambda}'\right\}, \tag{3.11}$$

where $\boldsymbol{\lambda} = (\delta, \boldsymbol{\theta}')'$, $\boldsymbol{\tau}_p$ *is the* p*-dimensional unit vector, and* $rank(\boldsymbol{\lambda}\boldsymbol{\lambda}') = 1$. *The OLS estimators of* α_0 *and* $\boldsymbol{\beta}^*$, *denoted by* $\hat{\alpha}_{0T}$ *and* $\hat{\boldsymbol{\beta}}_T^*$, *are also* $\sqrt{T}$*-consistent, and have the mixture normal distributions, asymptotically. The covariance matrix for all the short-run parameters,* $\mathbf{h} = (\mathbf{f}', \boldsymbol{\phi})'$, *is asymptotically singular with rank equal to* $kq + 2$, *and can be consistently estimated in the usual way by*

$$\hat{V}(\hat{\mathbf{h}}_T) = \hat{\sigma}_{uT}^2 (\mathbf{P}_{G_T}' \mathbf{P}_{G_T})^{-1},$$

where $\mathbf{P}_{G_T} = (\mathbf{G}_T, \mathbf{Y}_T)$, *and* $\hat{\sigma}_{uT}^2 = T^{-1}(\mathbf{y}_T - \mathbf{P}_{G_T}\hat{\mathbf{h}}_T)'(\mathbf{y}_T - \mathbf{P}_{G_T}\hat{\mathbf{h}}_T)$.

From Theorem 3.1 we also find that $\sqrt{T}(\hat{\alpha}_{1T}-\alpha_1)$ and $\sqrt{T}(\hat{\boldsymbol{\beta}}_T-\boldsymbol{\beta})$ are asymptotically perfectly collinear with $\sqrt{T}(\hat{\boldsymbol{\phi}}_T-\boldsymbol{\phi})$; that is,

$$\sqrt{T}\left\{(\hat{\mathbf{c}}_T-\mathbf{c})+\boldsymbol{\lambda}[\hat{\phi}_T(1)-\phi(1)]\right\}=o_p(1). \tag{3.12}$$

where $\hat{\phi}_T(1)=1-\sum_{i=1}^{p}\hat{\phi}_{iT}$. It is also straightforward to show that

$$\hat{\boldsymbol{\lambda}}_T-\boldsymbol{\lambda}=\frac{(\hat{\mathbf{c}}_T-\mathbf{c})+\boldsymbol{\lambda}[\hat{\phi}_T(1)-\phi(1)]}{\hat{\phi}_T(1)}. \tag{3.13}$$

Using Theorem 3.1, and results (3.12) and (3.13), we have:

Theorem 3.2. *Under the assumptions (A1)′ and (A2) - (A5), the OLS estimators of the long-run parameters, $\hat{\boldsymbol{\lambda}}_T=(\hat{\delta}_T,\hat{\boldsymbol{\theta}}'_T)'=\hat{\mathbf{c}}_T/\hat{\phi}_T(1)$ in (3.9), converge to their true values at faster rates than the estimators of the associated short-run parameters, and follow the mixture normal distribution asymptotically. Therefore,*

$$\mathbf{Q}^{\frac{1}{2}}_{\tilde{S}_T}\mathbf{D}^{-1}_{S_T}(\hat{\boldsymbol{\lambda}}_T-\boldsymbol{\lambda})\overset{a}{\sim}N\left\{\mathbf{0},\frac{\sigma_u^2}{[\phi(1)]^2}\mathbf{I}_{k+1}\right\}, \tag{3.14}$$

where $\mathbf{Q}_{\tilde{S}_T}$ and $\mathbf{D}_{S_T}$ are as defined in Theorem 2.2.

Comparing Theorems 2.2 and 3.2, we find that the presence of the I(0) stationary regressors in (3.9) (i.e., additional lagged changes in y_t and the lagged changes in $\mathbf{x}_t$ which are introduced to deal with the residual serial correlation problem) does not affect the asymptotic properties of the OLS estimator of the long run coefficients, δ and $\boldsymbol{\theta}$. Therefore, inferences concerning the long-run parameters can be based on the same standard tests as given by (2.12) and (2.13). In this more general case, however, the expression for the asymptotic variance of $\hat{\boldsymbol{\lambda}}_T$ is still given by (2.11), but with $\sigma_u^2/(1-\phi)^2$ replaced by the more general expression, $\sigma_u^2/[\phi(1)]^2$.

We now relax assumption (A3) and allow for the possibility of endogenous regressors, but confine our attention to the case where $\Delta\mathbf{x}_t$ can be represented by a finite order vector AR(s) process,[9]

$$\mathbf{P}(L)\Delta\mathbf{x}_t=\boldsymbol{\varepsilon}_t, \tag{3.15}$$

where $\mathbf{P}(L)=\mathbf{I}_k-\sum_{i=1}^{s}\mathbf{P}_i$, and $\mathbf{P}_i$, $i=1,...,s$, are the $k\times k$ coefficient matrices such that the vector autoregressive process in $\Delta\mathbf{x}_t$ is stable. Here $\boldsymbol{\varepsilon}_t$ are assumed

[9] Our analysis can also allow for the inclusion of lagged Δy's and a drift term in (3.15) without affecting the results presented below. On this see Boswijk (1995) and Pesaran, Shin and Smith (1996).

to be serially uncorrelated, but possibly contemporaneously correlated with u_t; namely, we assume that $\boldsymbol{\zeta}_t = (u_t, \boldsymbol{\varepsilon}_t')'$ follows the multivariate *iid* process with mean zero and the covariance matrix,

$$\boldsymbol{\Sigma}_{\zeta\zeta} = \begin{bmatrix} \sigma_u^2 & \boldsymbol{\Sigma}_{u\varepsilon} \\ \boldsymbol{\Sigma}_{\varepsilon u} & \boldsymbol{\Sigma}_{\varepsilon\varepsilon} \end{bmatrix}. \tag{3.16}$$

We will, however, continue to assume that $Cov(u_{t-j}, \boldsymbol{\varepsilon}_{t-i}) = 0$ for $i \neq j$. Notice that despite this assumption the model is still general enough to allow not only for the contemporaneous but also for cross-autocorrelations between u_t and $\Delta\mathbf{x}_t$. With assumption (A3) relaxed, the OLS estimators in (3.1) are no longer consistent. To correct for the endogeneity of $\mathbf{x}_t$, we model the contemporaneous correlation between u_t and $\boldsymbol{\varepsilon}_t$ by the linear regression of u_t on $\boldsymbol{\varepsilon}_t$

$$u_t = \mathbf{d}'\boldsymbol{\varepsilon}_t + \eta_t, \tag{3.17}$$

where using (3.16) we have $\mathbf{d} = \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\boldsymbol{\Sigma}_{u\varepsilon}'$, and $\boldsymbol{\varepsilon}_t$ is strictly exogenous with respect to η_t.[10] Substituting (3.15) in (3.17) we obtain:

$$u_t = \mathbf{d}'\mathbf{P}(L)\Delta\mathbf{x}_t + \eta_t, \tag{3.18}$$

where $\Delta\mathbf{x}_{t-i}$'s, $i = 0, ..., s$, are also strictly exogenous with respect to η_t. The parametric correction for the endogenous regressors is then equivalent to extending the ARDL(p, q) model (3.2) to the more general ARDL(p, m) specification,

$$\phi(L)y_t = \alpha_0 + \alpha_1 t + \boldsymbol{\beta}'\mathbf{x}_t + \sum_{j=0}^{m-1} \boldsymbol{\pi}_j'\Delta\mathbf{x}_{t-j} + \eta_t, \tag{3.19}$$

where $m = max(q, s+1)$, and $\boldsymbol{\pi}_i = \boldsymbol{\beta}_i^* - \mathbf{P}_i'\mathbf{d}$, $i = 0, 1, 2, ..., m-1$, $\mathbf{P}_0 = \mathbf{I}_k$, $\boldsymbol{\beta}_i^* = 0$ for $i \geq q$, and $\mathbf{P}_i = 0$ for $i \geq s$.

We now replace assumption (A3) by

(A3)$'$ The scalar disturbance η_t in (3.19) is $iid(0, \sigma_\eta^2)$, and $\Delta\mathbf{x}_t$ follows the general stationary process given by (3.15). Furthermore, η_t and $\boldsymbol{\varepsilon}_t$ are uncorrelated such that $\mathbf{x}_t$ and $\Delta\mathbf{x}_{t-j}$'s $j = 0, ..., m-1$, are strictly exogenous with respect to η_t in the ARDL(p, m) model (3.19).

There are two main differences between the ARDL models defined by (3.2) and (3.19). Firstly, the order of lagged $\Delta\mathbf{x}_t$'s in the two models can differ, and secondly, the coefficients on $\Delta\mathbf{x}_t$'s and their lagged values have different interpretations. Although this alters the dynamic structure of the model, the basic framework for estimating the long-run parameters and carrying out statistical inference on them is the same as before.

[10]The relation (3.17) will be exact when the joint distribution of u_t and $\boldsymbol{\varepsilon}_t$ is normal.

Theorem 3.3. *Under the assumptions (A3)′, (A4) and (A5), the OLS estimators of the short-run parameters in (3.19), α_0, α_1, $\boldsymbol{\beta}$, $\phi_1, ..., \phi_p$, $\boldsymbol{\pi}_0, ..., \boldsymbol{\pi}_{m-1}$, are $\sqrt{T}$-consistent, and asymptotically have the (mixture) normal distributions. Furthermore, $\sqrt{T}(\hat{\mathbf{c}}_T - \mathbf{c})$ is asymptotically perfectly collinear with $\sqrt{T}\left[\hat{\phi}_T(1) - \phi(1)\right]$, where $\mathbf{c} = (\alpha_1, \boldsymbol{\beta}')'$ and $\phi(1) = 1 - \sum_{i=1}^{p} \phi_i$, such that the covariance matrix for the estimators of the short-run parameters is asymptotically singular with rank equal to $km + 2$. The asymptotic distribution of the OLS estimators of the long-run parameters, $\hat{\boldsymbol{\lambda}}_T = (\hat{\delta}_{T,}\hat{\boldsymbol{\theta}}_T')' = \hat{\mathbf{c}}_T/\hat{\boldsymbol{\phi}}_T(1)$ in (3.19), are mixture normal and therefore,*

$$\mathbf{Q}_{\tilde{S}_T}^{\frac{1}{2}} \mathbf{D}_{S_T}^{-1}(\hat{\boldsymbol{\lambda}}_T - \boldsymbol{\lambda}) \overset{a}{\sim} N\left\{\mathbf{0}, \frac{\sigma_\eta^2}{[\phi(1)]^2}\mathbf{I}_{k+1}\right\}, \tag{3.20}$$

where σ_η^2 is the variance of η_t in (3.19), and $\mathbf{Q}_{\tilde{S}_T}$ and $\mathbf{D}_{S_T}$ are as defined in Theorem 2.2.

There are no fundamental differences between the results of Theorems 2.2, 3.2 and 3.3, as far as the estimators of the log-run parameters are concerned. A comparison of (2.11), (3.14) and (3.20) shows that the asymptotic distributions of the estimators of the long-run parameters, $\hat{\boldsymbol{\lambda}}_T$, under various assumptions discussed above differ only by a scalar coefficient.

In sum, in the context of the ARDL model inference on the long run parameters, δ and $\boldsymbol{\theta}$, is quite simple and requires *a priori* knowledge or estimation of the orders of the extended ARDL(p, m) model. Appropriate modification of the orders of the ARDL model is sufficient to simultaneously correct for the residual serial correlation and the problem of endogenous regressors. Variances of the OLS estimators of the long-run coefficients can then be consistently estimated using either (3.20), or by means of the Δ-method applied directly to the long-run estimators. Alternatively, one could compute the estimates of the long-run coefficients and their associated standard errors using Bewley's (1979) regression procedure. Bewley's method involves rewriting (3.19) as

$$\phi(L)y_t = \frac{\alpha_0}{\phi(1)} + \delta t + \boldsymbol{\theta}'\mathbf{x}_t + \frac{1}{\phi(1)}\sum_{j=0}^{m-1}\boldsymbol{\pi}_j'\Delta\mathbf{x}_{t-j} - \frac{1}{\phi(1)}\sum_{j=0}^{p-1}\phi_j^*\Delta y_{t-j} + \frac{\eta_t}{\phi(1)}, \tag{3.21}$$

and then estimating it by the instrumental variable method using (1, t, $\mathbf{x}_t$, $\Delta\mathbf{x}_t$, $\Delta\mathbf{x}_{t-1}, ..., \Delta\mathbf{x}_{t-m+1}, y_{t-1}, y_{t-2}, ..., y_{t-p}$) as instruments. It is easy to show that the IV estimators of δ and $\boldsymbol{\theta}$ obtained using (3.21) are *numerically* identical to the OLS estimators of δ and $\boldsymbol{\theta}$ based on the ARDL model (3.19), and that the standard errors of the IV estimators from the Bewley's regression are *numerically* identical to the standard errors of the OLS estimators of δ and $\boldsymbol{\theta}$ obtained using the Δ-method. (See, for example, Bardsen (1989).) The main attraction of the Bewley's regression procedure lies in its possible computational convenience as compared to the direct OLS estimation of (3.19) and computation of the associated standard errors by the Δ-method.[11]

Finally, we note in passing that the results developed in this section also apply to the case where the underlying regressors, $\mathbf{x}_t$, given by (3.15), are I(0). (See footnote 7 and the Monte Carlo simulation results in Section 5.)

4. A Comparison of ARDL and Phillips-Hansen Procedures

Here we focus on the case where there exists a unique cointegrating relation between I(1) variables, y_t and $\mathbf{x}_t$, possibly with a deterministic trend. The case where there are multiple cointegrating relations among I(1) variables presents additional difficulties and will not be discussed in this paper. (See Pesaran and Shin (1995), and the references cited therein).

Consider the following cointegrating relation

$$y_t = \mu + \delta t + \boldsymbol{\theta}' \mathbf{x}_t + v_t, \tag{4.1}$$

$$\Delta \mathbf{x}_t = \mathbf{e}_t. \tag{4.2}$$

Although the OLS estimator of $\boldsymbol{\theta}$ is shown to be T-consistent, (see Stock (1987)), it has also been found that the finite sample behavior of the OLS estimator is generally very poor (see, for example, Banerjee *et. al.* (1986)). Especially, in the presence of non-zero correlation between v_t and $\mathbf{e}_t$, OLS estimators of $\boldsymbol{\theta}$ in (4.1) are often heavily biased in finite samples, and inferences based on them are invalid because of the dependence of the limiting distribution of the OLS estimators on nuisance parameters. For details see Phillips and Loretan (1991).

Broadly speaking, there are two basic approaches to cointegration analysis: Johansen's (1991) maximum likelihood approach, and Phillips-Hansen's (1990, PH) fully modified OLS procedure.[12] The ARDL approach to cointegration analysis advanced in this paper is directly comparable to the PH procedure, and we shall, therefore concentrate on this method. PH assume that v_t and $\mathbf{e}_t$ in (4.1) and (4.2) follow the general correlated linear stationary processes:[13]

$$v_t = A_1(L)u_t, \ \mathbf{e}_t = \mathbf{A}_2(L)\boldsymbol{\varepsilon}_t, \tag{4.3}$$

[11] For a computer implementation of the ARDL approach using the Δ-method see Pesaran and Pesaran (1997).

[12] There are also other related procedures such as the original two-step method of Engle and Granger (1987), the leads and lags estimation procedure suggested by Saikkonen (1991) and Stock and Watson (1993), and the canonical method by Park (1992).

[13] For more details see Phillips and Solo (1992).

where $\boldsymbol{\zeta}_t = (u_t, \boldsymbol{\varepsilon}_t')'$ are serially uncorrelated random variables with zero means and a constant variance matrix given by (3.16). Assuming $A_1(L)$ and $\mathbf{A}_2(L)$ are invertible, (4.1) can be approximated as an ARDL specification by truncating the order of the infinite order lag polynomials $[A_1(L)]^{-1}$ and $[\mathbf{A}_2(L)]^{-1}$ such that $\phi(L) \approx [A_1(L)]^{-1}$ and $\mathbf{P}(L) \approx [\mathbf{A}_2(L)]^{-1}$, where the orders of the lag polynomials $\phi(L)$ and $\mathbf{P}(L)$ are denoted by p and s, respectively. Then we obtain the approximate finite-dimensional ARDL(p, m) specification,

$$\phi(L)y_t = \{\phi(1)\mu + \delta\phi'(1)\} + \delta\phi(1)t + \phi(L)\boldsymbol{\theta}'\mathbf{x}_t + \boldsymbol{\Sigma}_{u\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{P}(L)\Delta\mathbf{x}_t + \eta_t, \quad (4.4)$$

where $\phi'(1) = -\sum_{i=1}^{p} i\phi_i$, $m = max(p, s+1)$, and by construction $\mathbf{x}_t$ (and $\Delta\mathbf{x}_t$'s) are uncorrelated with η_t.[14] Notice that (4.4) is of the same form as (3.19), with the following relations among their parameters: $\alpha_0 = \phi(1)\mu + \delta\phi'(1)$, $\alpha_1 = \delta\phi(1)$, $\boldsymbol{\beta} = \phi(1)\boldsymbol{\theta}$, $\boldsymbol{\pi}'(L) = \phi^*(L)\boldsymbol{\theta}' + \boldsymbol{\Sigma}_{u\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{P}(L)$, where $\phi^*(L)$ is defined by $\phi(L) = \phi(1) + (1-L)\phi^*(L)$. Therefore, the ARDL specification (4.4) and the static cointegrating formulation, (4.1) and (4.2), represent alternative ways of modelling the serial correlation in v_t's and the endogeneity of $\mathbf{x}_t$.

Here we examine the PH estimation procedure in the context of the ARDL approximation for the y_t process given by (4.4). Assuming that $\boldsymbol{\xi}_t = (v_t, \mathbf{e}_t')'$ in (4.1) and (4.2) satisfy the multivariate invariance principle, the long-run variance matrix of $\boldsymbol{\xi}_t$ is given by[15]

$$\boldsymbol{\Omega}_\xi = \underset{T\to\infty}{\text{Plim}} \left\{ T^{-1}\sum_{t=1}^{T} \boldsymbol{\xi}_t\boldsymbol{\xi}_t' + T^{-1}\sum_{j=1}^{\ell}\left[\sum_{t=j+1}^{T}\boldsymbol{\xi}_t\boldsymbol{\xi}_{t-j}' + \sum_{t=j+1}^{T}\boldsymbol{\xi}_{t-j}\boldsymbol{\xi}_t'\right]\right\}, \quad (4.5)$$

where the lag truncation parameter ℓ increases with T, such that $\ell/T \to 0$, as $T \to \infty$. We also define

$$\boldsymbol{\Delta}_\xi = \underset{T\to\infty}{\text{Plim}}\, T^{-1}\left\{\sum_{t=1}^{T}\boldsymbol{\xi}_t\boldsymbol{\xi}_t' + \sum_{j=1}^{\ell}\sum_{t=j+1}^{T}\boldsymbol{\xi}_t\boldsymbol{\xi}_{t-j}'\right\}, \quad (4.6)$$

and partition $\boldsymbol{\Omega}_\xi$ and $\boldsymbol{\Delta}_\xi$ conformably to $\boldsymbol{\xi}_t = (v_t, \mathbf{e}_t')'$,

$$\boldsymbol{\Omega}_\xi = \begin{bmatrix} \omega_{vv} & \boldsymbol{\Omega}_{ve} \\ \boldsymbol{\Omega}_{ev} & \boldsymbol{\Omega}_{ee} \end{bmatrix}, \; \boldsymbol{\Delta}_\xi = \begin{bmatrix} \Delta_{vv} & \boldsymbol{\Delta}_{ve} \\ \boldsymbol{\Delta}_{ev} & \boldsymbol{\Delta}_{ee} \end{bmatrix}.$$

Although the use of the consistent estimator of the long-run variance matrix may solve the serial correlation problem of v_t, this does not address the endogeneity

[14] As before, $\eta_t = u_t - \boldsymbol{\Sigma}_{u\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1}\boldsymbol{\varepsilon}_t$.

[15] The random sequence $\{\boldsymbol{\xi}_t\}$ is said to satisfy the multivariate invariance principle if it is strictly stationary and ergodic with zero mean, finite variances, and spectral density matrix $f_{\xi\xi}(\omega) > 0$. See Phillips and Durlauf (1986) for details.

problem. To deal with the cross-correlations between v_t and current and lagged values of $\mathbf{e}_t$, PH consider the modified error process, denoted by v_t^+, which is obtained from the regression of v_t on $\mathbf{e}_t$,

$$v_t^+ = v_t - \boldsymbol{\Omega}_{ve}\boldsymbol{\Omega}_{ee}^{-1}\mathbf{e}_t, \tag{4.7}$$

and v_t^+ is not correlated with $\mathbf{e}_t$ by construction. Then, the long-run variance matrix of $\boldsymbol{\xi}_t^+ = (v_t^+, \mathbf{e}_t')'$, denoted by $\boldsymbol{\Omega}_\xi^+$, is block diagonal; that is, $\boldsymbol{\Omega}_\xi^+ = diag(\omega_{v\cdot e}, \boldsymbol{\Omega}_{ee})$, where

$$\omega_{v\cdot e} = \omega_{vv} - \boldsymbol{\Omega}_{ve}\boldsymbol{\Omega}_{ee}^{-1}\boldsymbol{\Omega}_{ev}, \tag{4.8}$$

is the conditional long-run variance of v_t given $\mathbf{e}_t$. Combining (4.7) with (4.1) we have the modified "static" cointegrating relation,

$$y_t^+ = \mu + \delta t + \boldsymbol{\theta}'\mathbf{x}_t + v_t^+, \tag{4.9}$$

where $y_t^+ = y_t - \boldsymbol{\Omega}_{ve}\boldsymbol{\Omega}_{ee}^{-1}\Delta\mathbf{x}_t$. There is still a bias term remaining in (4.9) because of the correlation between $\mathbf{x}_t$ and current and lagged values of v_t^+, which is given by $\boldsymbol{\Delta}_{ev}^+ = \boldsymbol{\Delta}_{ev} - \boldsymbol{\Delta}_{ee}\boldsymbol{\Omega}_{ee}^{-1}\boldsymbol{\Omega}_{ev}$. Removing this bias leads to the Phillips-Hansen fully-modified OLS estimators,

$$\begin{bmatrix} \hat{\mu}_T^+ \\ \hat{\delta}_T^+ \\ \hat{\boldsymbol{\theta}}_T^+ \end{bmatrix} = (\mathbf{Z}_T'\mathbf{Z}_T)^{-1}\left\{\mathbf{Z}_T'\hat{\mathbf{y}}_T^+ - \begin{bmatrix} 0 \\ 0 \\ \boldsymbol{\tau}_k \end{bmatrix}\right\} T\hat{\boldsymbol{\Delta}}_{ev}^+, \tag{4.10}$$

where $\mathbf{Z}_T = (\boldsymbol{\tau}_T, \mathbf{t}_T, \mathbf{X}_T)$, $\boldsymbol{\tau}_k$ is the k-dimensional column unit vector, and $\hat{\mathbf{y}}_T^+$ and $\hat{\boldsymbol{\Delta}}_{ev}^+$ are consistent estimators of y_t^+ and $\boldsymbol{\Delta}_{ev}^+$, respectively.

Since the asymptotic distribution of the PH estimators of the coefficients on t and $\mathbf{x}_t$ (standardized by $T^{\frac{3}{2}}$ and T, respectively) is (mixture) normal, we have

$$\mathbf{Q}_{\tilde{S}_T}^{\frac{1}{2}}\mathbf{D}_{S_T}^{-1}(\hat{\boldsymbol{\lambda}}_T^+ - \boldsymbol{\lambda}) \overset{a}{\sim} N\left\{0,\ \omega_{v\cdot e}\mathbf{I}_{k+1}\right\}, \tag{4.11}$$

where $\hat{\boldsymbol{\lambda}}_T^+ = (\hat{\delta}_T^+, \hat{\boldsymbol{\theta}}_T^{+\prime})'$. This is directly comparable to the asymptotic result in (3.20) obtained using the ARDL estimation procedure. First, we find that the estimators of the long run parameters obtained using both the ARDL and the PH estimation procedures have the mixture normal distributions asymptotically, and standard inferences on $\boldsymbol{\theta}$ using the Wald test are therefore asymptotically valid. The main difference between the ARDL-based estimators and the fully-modified OLS estimators lies in the computation of the long-run variance of the disturbances in the cointegrating regression. In the case of the ARDL estimation procedure the long run variance is given by $\sigma_\eta^2/[\phi(1)]^2$, while in the case of the PH estimation procedure the long run variance is given by $\omega_{v\cdot e}$. But as Theorem 8 below shows, $\sigma_\eta^2/[\phi(1)]^2$ and $\omega_{v\cdot e}$ are identical for the ARDL specification (3.19) (or (4.4)).

Theorem 4.1. *In the context of the ARDL specification (3.19) or (4.4), the long-run variance of the Phillips-Hansen modified error process,* v_t^+ *in (4.9) (denoted by* $\omega_{v\cdot e}$*) is equal to* $\sigma_\eta^2/[\phi(1)]^2$*, which is the spectral density at zero frequency of* $[\phi(L)]^{-1}\eta_t$ *in (3.19).*

5. Finite Sample Simulation Results

In this section, using Monte Carlo techniques, we compare finite sample properties of the Phillips-Hansen fully-modified estimators of the long-run parameters with the ARDL-based estimators. In the case of the ARDL procedure we consider two different estimators of the variance of the long-run parameter, namely the asymptotic formula (2.19), which is valid only for I(1) regressors, and the Δ-method formula given by (2.20), which is valid more generally, irrespective of whether the regressors are I(1) or I(0). We also include the OLS estimators of the long-run parameters in the static cointegrating relation as a rather crude bench mark of interest.

We consider the following data generating process (DGP), where the observations on y_t and x_t are generated according to the finite-order ARDL (1,0) model:

$$y_t = \alpha + \phi y_{t-1} + \beta x_t + u_t, \tag{5.1}$$

$$x_t - \psi x_{t-1} = \rho\left(x_t - \psi x_{t-1}\right) + \varepsilon_t, \tag{5.2}$$

$t = 1, ..., T$, where (u_t, ε_t) are serially uncorrelated and are generated according to the following bivariate normal distribution:

$$\begin{pmatrix} u_t \\ \varepsilon_t \end{pmatrix} \sim N\left\{\mathbf{0}, \mathbf{\Omega} = \begin{pmatrix} 1 & \omega_{12} \\ \omega_{12} & 1 \end{pmatrix}\right\}. \tag{5.3}$$

We set $\alpha = 0$, $\beta = 1$, $\rho = 0.2$, and experiment with the following parameter values: $\phi = (0.2, 0.8)$, $\omega_{12} = (-0.5, 0.0, 0.5)$, and $T = (50, 100, 250)$.

We carry out two sets of experiments: In the first set (Experiments 1) we fix ψ at 1 and therefore, generate x_t as an I(1) process. In the second set (Experiments 2) we set ψ to 0.95 such that x_t is I(0) but with a high degree of persistence. It is worth noting that in general (irrespective of whether x_t is I(1) or I(0)), the long run parameter on x_t in (5.1) is given by

$$\theta = \frac{\beta + (1-\psi)\omega_{12}}{1-\phi},$$

and θ will be invariant to the parameters of the x_t process only if $\omega_{12} = 0$ (i.e., x_t is strictly exogenous in (5.1)) and/or when $\psi = 1$ (i.e., x_t is I(1)). For a more general treatment of this issue see Pesaran (1997).

Before discussing the simulation results, notice that when $\omega_{12} = 0$, the correct specification is the ARDL(1,0) model, and when $\omega_{12} \neq 0$, it is the ARDL(1,2) model. (See Section 3). But since in general the true order of the ARDL model is not known *a priori*, we estimated 30 different ARDL models, namely ARDL(p, m), $p = 1, 2, ..., 5$, $m = 0, 1, 2, ..., 5$, and used the Akaike Information Criterion (AIC), and the Schwarz Criterion (SC) to select the orders of the ARDL model before estimating the long-run coefficients and carrying out inferences. The estimates obtained by these two-step procedures will be referred to as ARDL-AIC, and ARDL-SC, respectively.

The simulation results are summarized in Tables 1a-1f and 2a-2f for Experiments 1 and 2, respectively. Summary statistics included in these tables are:

Bias $= \hat{\theta}_R - \theta_0$, where θ_0 is the true value of the long-run coefficient θ, $\hat{\theta}_R$ is the mean of the estimates of θ across replications, i.e., $\hat{\theta}_R = \sum_{i=1}^{R} \hat{\theta}_i / R$ and R is the number of replications,

STDE θ = Standard error of the estimator, $\hat{\theta}_i$, across replications,

RMSE = The root mean squared error of $\hat{\theta}_i$, $\left(\sqrt{R^{-1} \sum_{i=1}^{R} (\hat{\theta}_i - \theta_0)^2} \right)$,

Mean t = Average t-statistic for testing $\theta = \theta_0$ against $\theta \neq \theta_0$,

STD t = Standard deviations of the t-statistic for testing $\theta = \theta_0$ against $\theta \neq \theta_0$,

SIZE = Empirical size of the t-test of the null hypothesis $\theta = \theta_0$ against $\theta \neq \theta_0$,

POWER$^+$ = Empirical power of the t-test under the alternatives $\theta = 1.05\theta_0$,

POWER$^-$ = Empirical power of the t-test under the alternatives $\theta = 0.95\theta_0$.

The nominal size of the tests is set at 5 percent, and the number of replications at $R = 2,500$.[16]

Tables 1a-1f summarize the results for the correctly specified ARDL model (namely the ARDL(1,0) when $\omega_{12} = 0$, and the ARDL(1,2) for $\omega_{12} \neq 0$), the estimates based on ARDL-AIC and the ARDL-SC procedures, and the Phillips-Hansen fully modified estimators based on the Bartlett's window for window sizes 0, 5, 10, 20 and 40, which are reported under PH(0), PH(5), etc.

In the case where $\omega_{12} = 0$, the bias of the ARDL estimators is much smaller than that of the PH estimators. The extent of the bias crucially depends on the value of ϕ, and not surprisingly increases as ϕ is increased from 0.2 in Table 1a

[16] In a very small number of replications $\phi(1)$ was estimated to be in excess of 0.99. These cases are not included in the summary results.

to 0.8 in Table 1d. Also the RMSE's of the ARDL and the PH estimators are very similar when $\phi = 0.2$, but diverge considerably for $\phi = 0.8$. As can be seen from Table 1d, for $T = 50$, the RMSE of the ARDL estimators is about one-third of the RMSE of the PH estimators. The empirical sizes of the ARDL procedure are much more satisfactory than the ones obtained using the PH fully modified estimators. When $\omega_{12} = 0$, the sizes of the tests based on the ARDL estimators are generally reasonable and much nearer to their nominal size of 5 percent, than the sizes of tests based on the PH estimators.

Empirical sizes of the tests based on the ARDL estimators computed using the Δ-method tend to be much closer to their nominal values, than those computed using the asymptotic formula. This is particularly so when T is small. Therefore, in what follows, we shall focus on the ARDL test statistics that are computed using the Δ-method.

Another general feature of the simulation results is the slight superiority of the ARDL-SC method over the ARDL-AIC procedure; which is in accordance with the fact that the SC is a consistent model selection criterion, while the AIC is not. (See, for example, Lütkepohl (1991, Chapter 4)).

Finally, there is a clear tendency for the tests based on the PH method to over-reject in small samples, and the extent of this over-rejection increases with ϕ, and declines only slowly with the sample size, T. For example, for $\phi = 0.8$ and $T = 100$, the empirical sizes of the t-tests based on the PH method exceed 41 percent for all the five window sizes, and even for $T = 250$ do not fall below 20 percent. (See the column headed "SIZE" in Table 1d). By contrast the size of the test based on the Δ-method in Table 1d is reasonable even for $T = 50$. For the correct ARDL(1,0) specification, the size of the test based on the Δ-method is 7.2 percent and increases to 12.8 and 8.6 percents for the ARDL-AIC and the ARDL-SC procedures, respectively.

Similar results are obtained in the case where $\omega_{12} = 0.5$, and hence x_t and u_t are contemporaneously correlated. The ARDL estimators are now substantially less biased than the PH estimators. (See the column headed "BIAS" in Table 1e). Once again the performance of the PH estimators improves with the sample size, but very slowly. For $T = 250$, the bias of the PH estimators for the most favorable window size is still around -0.14, but the biases of the ARDL estimators lie between -0.0017 and 0.0024. The size performance of the two test procedures also closely mirrors these differences in the degree of biases of the estimators. The empirical size of the tests based on the PH method ranges between 60 to 85 percent for $T = 50$, and falls to around 21 percent for $T = 250$ and a window size of 20. The size of the tests based on the ARDL procedure, when the Δ-method is used to compute the variances, is at most 13 percent for $T = 50$, and lies in the range 5.2 to 7.7 percent when T is increased to 250. (See Table 1e).

Due to the large size distortions of the PH procedure, the results presented in Tables 1a-1f do not allow proper comparisons of the power properties of the two test procedures. But for $T = 250$ where the size distortion of the PH test is not too excessive, the ARDL procedure consistently outperforms the PH method. For example, in the case of $\phi = 0.8$, $\omega_{12} = 0.5$, $\theta = 5$, and $T = 250$, the power of the ARDL procedure in rejecting the false null hypothesis, $\theta = 0.95\theta_0$, is consistently above 98 percent while the power of the PH method is at most 62 percent even though its associated size is 85 percent! There seems also to be a tendency for the power function of the ARDL procedure in the case where $\omega_{12} \neq 0$ and T small to be asymmetric around $\theta = \theta_0$; showing a higher power for the alternatives exceeding θ_0 as compared to the alternatives falling below θ_0.

The results for Experiments 2 with an I(0) regressor are summarized in Tables 2a-2f. These results are very similar to those obtained for Experiments 1. The overall performances of the ARDL-based methods with variances estimated using the Δ-method are satisfactory for most cases, though slightly worse than those obtained for Experiments 1. (In particular, the biases are slightly larger and the tests are less powerful.) But, the performance of the PH estimators are still very poor, especially when T is small.

Overall, the simulation results show that the ARDL-based estimation procedure based on the Δ-method developed in the paper can be reliably used in small samples to estimate and test hypotheses on the long-run coefficients in both cases where the underlying regressors are I(1) or I(0). This is an important finding since the ARDL approach can avoid the pretesting problem implicitly involved in the cointegration analysis of the long-run relationships. (Also see Cavanaugh *et. al.* (1995) and Pesaran (1997).)

Before concluding this section, we note that the comparison of the small sample performance of the ARDL-based and the PH estimators is not comprehensive in the sense that the data generating process we have used is biased in favor of the ARDL procedure (see Inder (1993)). In this regard, it is more appropriate to consider the relative performances of the ARDL and the PH estimators using more general DGP's, such as (4.1) and (4.2), that can allow for moving average error processes. In the working paper version of this paper we also considered Monte Carlo experiments using (4.1) and (4.2) as data generating processes. In one set of experiments (called $DGP2$) we used first-order bivariate vector moving-average processes to generate the errors, v_t and e_t, and in another set of experiments (called $DGP3$) we generated v_t and e_t according to first-order vector autoregressive processes. Neither of these DGP's allows transformations of the model so that x_t could become strictly exogenous with respect to the disturbances of the augmented ARDL model. We found that the simulation results based on these DGP's are less clear-cut, but the ARDL-based estimator using the Δ-method still outperforms the PH estimator in most experiments, especially for small T. Broadly speaking, the relative small sample performance of the two estimators seems to depend on the signal-to-noise ratio, $Var(e_t)/Var(v_t)$, with the ARDL approach dominating the PH method when this ratio is low, and *vice versa*. This is clearly an area for further research.[17]

6. Concluding Remarks

The theoretical analysis and the Monte Carlo results presented in this paper provide strong evidence in favor of a rehabilitation of the traditional ARDL approach to time series econometric modelling. The focus of this paper, however, has been exclusively on single equation estimation techniques and the important issue of system estimation is not addressed here. Such an analysis inevitably involves the problem of identification of short-run and long-run relations and demands a structural approach to the analysis of econometric models. The problem of long-run structural modelling in the context of an unrestricted VAR model has been addressed elsewhere. (See, for example, Johansen (1991), Phillips (1991) and Pesaran and Shin (1995)). An alternative procedure, which takes us back to the Cowles Commission approach, would be to extend the ARDL methodology advanced in this paper to systems of equations subject to short-run and/or long-run identifying restrictions. (See, for example, Boswijk (1995) and Hsiao (1995).) We hope to pursue this line of research in the future; thus establishing a closer link between the recent cointegration analysis and the traditional simultaneous equations econometric methodology.

[17] We are grateful to Peter Boswijk and an anonymous referee for drawing our attention to this point.

Appendix: Mathematical Proofs

For notational convenience we use "$\stackrel{p}{\to}$", "$\Rightarrow$" and "$\stackrel{a}{\sim}$" to signify the convergence in probability, the weak convergence in probability measure, and the asymptotic equality in distribution. All sums are over $t = 1, 2, ..., T$.

In the case where the regressors are stationary the usual method of deriving the asymptotic distribution of the OLS estimators of the short-run parameters in, for example, (2.1), would be to apply the Slutsky's theorem to $(\mathbf{P}'_{Z_T}\mathbf{P}_{Z_T})^{-1}$ and $\mathbf{P}'_{Z_T}\mathbf{u}_T$, separately, where $\mathbf{P}_{Z_T} = (\boldsymbol{\tau}_T, \mathbf{t}_T, \mathbf{X}_T, \mathbf{y}_{T-1})$, after appropriately scaling it by the sample size. (The appropriate scaling of $\mathbf{P}'_{Z_T}\mathbf{P}_{Z_T}$ in this case is given by $\mathbf{D}_{P_T}\mathbf{P}_{Z_T}\mathbf{P}'_{Z_T}\mathbf{D}_{P_T}$ where $\mathbf{D}_{P_T} = Diag(T^{-\frac{1}{2}}, T^{-\frac{3}{2}}, T^{-1}\mathbf{I}_k, T^{-1})$.) This procedure cannot, however, be applied to dynamic time series models with trended regressors (irrespective of whether the trends are stochastic or deterministic), because $\mathbf{P}'_{Z_T}\mathbf{P}_{Z_T}$ does not converge to a non-singular matrix even if the individual elements of $\mathbf{P}'_{Z_T}\mathbf{P}_{Z_T}$ are appropriately scaled by the sample size.

In what follows the asymptotic theory will be developed using the partitioned regression techniques and then writing individual elements of the OLS estimators of the short-run parameters as ratios of random variables, thus avoiding the need to apply the Slutsky's theorem to $(\mathbf{P}'_{Z_T}\mathbf{P}_{Z_T})^{-1}$ directly.

Since Theorems 2.1 - 2.4 are special cases of Theorems 3.1 and 3.2, and can be proved in a similar manner, we omit their proofs to save space.

Proof of Theorem 3.1.

Before deriving the asymptotic distributions of the OLS estimators of the short run parameters in (3.9) we derive some preliminary results. Define

$$\mathbf{q}_{K_T u_T} = T^{-\frac{1}{2}}\mathbf{K}'_T\mathbf{u}_T,\ \mathbf{Q}_{K_T} = T^{-1}\mathbf{K}'_T\mathbf{K}_T,$$

$$\mathbf{q}_{G_T u_T} = \mathbf{D}_{G_T}\mathbf{G}'_T\mathbf{u}_T = \begin{bmatrix} \mathbf{D}_{Z_T}\mathbf{Z}'_T\mathbf{u}_T \\ T^{-\frac{1}{2}}\mathbf{W}'_T\mathbf{u}_T \end{bmatrix} = \begin{bmatrix} \mathbf{q}_{Z_T u_T} \\ \mathbf{q}_{W_T u_T} \end{bmatrix},$$

$$\mathbf{q}_{G_T K_T} = \mathbf{D}_{G_T}\mathbf{G}'_T\mathbf{K}_T = \begin{bmatrix} \mathbf{D}_{Z_T}\mathbf{Z}'_T\mathbf{K}_T \\ T^{-\frac{1}{2}}\mathbf{W}'_T\mathbf{K}_T \end{bmatrix} = \begin{bmatrix} \mathbf{q}_{Z_T K_T} \\ \mathbf{q}_{W_T K_T} \end{bmatrix},$$

$$\mathbf{Q}_{G_T} = \mathbf{D}_{G_T}\mathbf{G}'_T\mathbf{G}_T\mathbf{D}_{G_T} = \begin{bmatrix} \mathbf{D}_{Z_T}\mathbf{Z}'_T\mathbf{Z}_T\mathbf{D}_{Z_T} & T^{-\frac{1}{2}}\mathbf{D}_{Z_T}\mathbf{Z}'_T\mathbf{W}_T \\ T^{-\frac{1}{2}}\mathbf{W}'_T\mathbf{Z}_T\mathbf{D}_{Z_T} & T^{-1}\mathbf{W}'_T\mathbf{W}_T \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{Z_T} & \mathbf{Q}_{Z_T W_T} \\ \mathbf{Q}'_{Z_T W_T} & \mathbf{Q}_{W_T} \end{bmatrix},$$

where $\mathbf{K}_T = (\boldsymbol{\kappa}_{1T}, \boldsymbol{\kappa}_{2T}, ..., \boldsymbol{\kappa}_{pT})$ with $\boldsymbol{\kappa}_{iT} = (\kappa_{i1}, \kappa_{i2}, ..., \kappa_{iT})'$ for $i = 1, ..., p$, $\mathbf{D}_{G_T} = Diag(T^{-\frac{1}{2}}, T^{-\frac{3}{2}}, T^{-1}\mathbf{I}_k, T^{-\frac{1}{2}}\mathbf{I}_{kq})$ and $\mathbf{D}_{Z_T} = Diag(T^{-\frac{1}{2}}, T^{-\frac{3}{2}}, T^{-1}\mathbf{I}_k)$. Using the results in Phillips and Durlauf (1986), it is easily seen that as $T \to \infty$,

$$\mathbf{q}_{K_T u_T} \stackrel{p}{\to} \mathbf{q}_{Ku},\ \mathbf{Q}_{K_T} \stackrel{p}{\to} \mathbf{Q}_{K}, \tag{A.1}$$

$$\mathbf{q}_{G_T u_T} \Rightarrow \mathbf{q}_{Gu} = \begin{bmatrix} \mathbf{q}_{Zu} \\ \mathbf{q}_{Wu} \end{bmatrix}, \; \mathbf{q}_{G_T K_T} \Rightarrow \mathbf{q}_{GK} = \begin{bmatrix} \mathbf{q}_{ZK} \\ \mathbf{q}_{WK} \end{bmatrix}, \tag{A.2}$$

$$\mathbf{Q}_{G_T} \Rightarrow \mathbf{Q}_G = \begin{bmatrix} \mathbf{Q}_Z & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_W \end{bmatrix}, \tag{A.3}$$

where $\mathbf{q}_{Ku}$, $\mathbf{q}_{Wu}$, $\mathbf{q}_{WK}$, $\mathbf{Q}_K$ and $\mathbf{Q}_W$ are (finite) probability limits of $\mathbf{q}_{K_T u_T}$, $\mathbf{q}_{W_T u_T}$, $\mathbf{q}_{W_T K_T}$, $\mathbf{Q}_{K_T}$ and $\mathbf{Q}_{W_T}$, respectively, and $\mathbf{q}_{Zu}$, $\mathbf{q}_{ZK}$ and $\mathbf{Q}_Z$ are functionals of Brownian motions given by

$$\mathbf{q}_{Zu} = \begin{bmatrix} B_u(1) \\ \int_0^1 r dB_u(r) \\ \int_0^1 \mathbf{B}_e'(r) dB_u(r) \end{bmatrix}, \; \mathbf{q}_{ZK} = \begin{bmatrix} \mathbf{B}_K(1) \\ \int_0^1 r d\mathbf{B}_K(r) \\ \int_0^1 \mathbf{B}_e'(r) d\mathbf{B}_K(r) \end{bmatrix},$$

$$\mathbf{Q}_Z = \begin{bmatrix} 1 & \frac{1}{2} & \int_0^1 \mathbf{B}_e(r)dr \\ \frac{1}{2} & \frac{1}{3} & \int_0^1 r\mathbf{B}_e(r)dr \\ \int_0^1 \mathbf{B}_e'(r)dr & \int_0^1 r\mathbf{B}_e'(r)dr & \int_0^1 \mathbf{B}_e'(r)\mathbf{B}_e(r)dr \end{bmatrix}.$$

$B_u(r)$ is the scalar Brownian motion process with variance equal to r times σ_u^2 (since u_t is not serially correlated), $\mathbf{B}_e(r)$ is a k-dimensional Brownian motion on $r \in [0,1]$ with variance equal to r times the long-run variance of $\mathbf{e}_t$, and $\mathbf{B}_K(r)$ is the p-dimensional Brownian motion on [0,1] with variance equal to r times the long run variance of $(\boldsymbol{\kappa}_{1T}, \boldsymbol{\kappa}_{2T}, ..., \boldsymbol{\kappa}_{pT})$. The long-run variance of a stochastic process is given by 2π multiplied by the spectral density of the process at zero frequency. Notice that $\mathbf{Q}_Z$ (or $\mathbf{Q}_G$) is of the full column rank by assumption (A4), and the elements in $\mathbf{Q}_Z$ involving $\mathbf{B}_e(r)$ are random even asymptotically.

Since $\boldsymbol{\kappa}_{1T}, \boldsymbol{\kappa}_{2T}, ..., \boldsymbol{\kappa}_{pT}$, and 1, t, $\mathbf{x}_t$, $\Delta\mathbf{x}_t$, $\Delta\mathbf{x}_{t-1}, ..., \Delta\mathbf{x}_{t-q+1}$ are all distributed independently of u_t such that $\mathbf{B}_K(r)$ and $\mathbf{B}_e(r)$ are independent of $B_u(r)$, it follows that

$$\mathbf{q}_{Ku} \overset{a}{\sim} N\left(\mathbf{0}, \sigma_u^2 \mathbf{Q}_\kappa\right), \; \mathbf{q}_{Gu} \overset{a}{\sim} MN\left(\mathbf{0}, \sigma_u^2 \mathbf{Q}_G\right), \tag{A.4}$$

where MN denotes the mixture normal distribution. For details concerning the theory of the mixture normal distribution see, for example, Phillips (1991). However, this (mixture) normality result does not hold in the case of $\mathbf{q}_{GK}$, because $\mathbf{x}_t$ and $\Delta\mathbf{x}_{t-i}$'s $(i = 0, ..., q-1)$ are correlated with $\boldsymbol{\kappa}_{it}$, $i = 1, ..., p$.

The OLS estimators of $\mathbf{f}$ and $\boldsymbol{\phi}$ in (3.9), denoted by $\hat{\mathbf{f}}_T$ and $\hat{\boldsymbol{\phi}}_T$, satisfy the relations,

$$\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi} = \left(\mathbf{Y}_T'\mathbf{M}_{G_T}\mathbf{Y}_T\right)^{-1}\left(\mathbf{Y}_T'\mathbf{M}_{G_T}\mathbf{u}_T\right), \tag{A.5}$$

$$\hat{\mathbf{f}}_T - \mathbf{f} = \left(\mathbf{G}_T'\mathbf{G}_T\right)^{-1}\left[\mathbf{G}_T'\mathbf{u}_T - \mathbf{G}_T'\mathbf{Y}_T\left(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}\right)\right], \tag{A.6}$$

where $\mathbf{M}_{G_T} = \mathbf{I}_T - \mathbf{G}_T(\mathbf{G}_T'\mathbf{G}_T)^{-1}\mathbf{G}_T'$ with $\mathbf{I}_T$ being the $T \times T$ identity matrix. Using (3.7), $\mathbf{Y}_T$ can be expressed as

$$\mathbf{Y}_T = \mathbf{G}_T\boldsymbol{\Gamma} + \mathbf{K}_T, \tag{A.7}$$

where

$$\mathbf{\Gamma} = \begin{bmatrix} \mu_1 & \mu_2 & \cdots & \mu_p \\ \delta & \delta & \cdots & \delta \\ \boldsymbol{\theta} & \boldsymbol{\theta} & \cdots & \boldsymbol{\theta} \\ \mathbf{g}_1 & \mathbf{g}_2 & \cdots & \mathbf{g}_p \end{bmatrix},$$

and $\mathbf{g}_i = (\mathbf{g}'_{i0}, \mathbf{g}'_{i1}, ..., \mathbf{g}'_{i,q-1})'$ is a $kq \times 1$ vector of parameters. Using (A.7) we have

$$\mathbf{Y}'_T \mathbf{M}_{G_T} \mathbf{Y}_T = \mathbf{K}'_T \mathbf{K}_T - \mathbf{K}'_T \mathbf{G}_T \left(\mathbf{G}'_T \mathbf{G}_T\right)^{-1} \mathbf{G}'_T \mathbf{K}_T,$$

$$\mathbf{Y}'_T \mathbf{M}_{G_T} \mathbf{u}_T = \mathbf{K}'_T \mathbf{u}_T - \mathbf{K}'_T \mathbf{G}_T \left(\mathbf{G}'_T \mathbf{G}_T\right)^{-1} \mathbf{G}'_T \mathbf{u}_T,$$

where we used $\mathbf{G}'_T \mathbf{M}_{G_T} = \mathbf{0}$. Using (A.1) - (A.3), it can be shown that as $T \to \infty$,

$$T^{-1} \left(\mathbf{Y}'_T \mathbf{M}_{G_T} \mathbf{Y}_T\right) = \mathbf{Q}_{K_T} + o_p(1) \xrightarrow{p} \mathbf{Q}_K, \tag{A.8}$$

$$T^{-\frac{1}{2}} \left(\mathbf{Y}'_T \mathbf{M}_{G_T} \mathbf{u}_T\right) = \mathbf{q}_{K_T u_T} + o_p(1) \xrightarrow{p} \mathbf{q}_{Ku}. \tag{A.9}$$

Multiplying (A.5) by $\sqrt{T}$, and using (A.8), (A.9) and (A.4), we obtain (3.10).

Next, substituting $\mathbf{Y}_T$ from (A.7) in (A.6), we obtain

$$\hat{\mathbf{f}}_T - \mathbf{f} = \left(\mathbf{G}'_T \mathbf{G}_T\right)^{-1} \mathbf{G}'_T \mathbf{u}_T - \mathbf{\Gamma} \left(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}\right) - \left(\mathbf{G}'_T \mathbf{G}_T\right)^{-1} \mathbf{G}'_T \mathbf{K}_T \left(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}\right). \tag{A.10}$$

Define

$$\mathbf{d}_T = \left(\hat{\mathbf{f}}_T - \mathbf{f}\right) + \mathbf{\Gamma} \left(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}\right). \tag{A.11}$$

Multiplying (A.11) by $\mathbf{D}_{G_T}^{-1}$, using (A.1) - (A.3) and (A.10), and applying the continuous mapping theorem (see, for example, Phillips and Durlauf (1986)), it follows that

$$\mathbf{D}_{G_T}^{-1} \mathbf{d}_T = \mathbf{Q}_{G_T}^{-1} \mathbf{q}_{G_T u_T} + o_p(1) \Rightarrow \mathbf{Q}_G^{-1} \mathbf{q}_{Gu}. \tag{A.12}$$

Since $\mathbf{q}_{Gu}$ is shown to be mixture normal in (A.4), hence

$$\mathbf{Q}_G^{-1} \mathbf{q}_{Gu} \overset{a}{\sim} MN\left(\mathbf{0}, \sigma_u^2 \mathbf{Q}_G^{-1}\right), \ \mathbf{Q}_{G_T}^{\frac{1}{2}} \mathbf{D}_{G_T}^{-1} \mathbf{d}_T \overset{a}{\sim} N\left(\mathbf{0}, \sigma_u^2 \mathbf{I}_{k+kq+2}\right).$$

Next, pre-multiplying (A.12) by the diagonal matrix, $Diag(1, T^{-1}, T^{-\frac{1}{2}}\mathbf{I}_k, \mathbf{I}_{kq})$, we have

$$\sqrt{T}\mathbf{d}_T = \begin{bmatrix} 1 & 0 & \mathbf{0} & \mathbf{0} \\ 0 & T^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & T^{-\frac{1}{2}}\mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 & \mathbf{I}_{kq} \end{bmatrix} \mathbf{Q}_{G_T}^{-1} \mathbf{q}_{G_T u_T} + o_p(1) \tag{A.13}$$

$$\Rightarrow \begin{bmatrix} 1 & 0 & \mathbf{0} & \mathbf{0} \\ 0 & 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{kq} \end{bmatrix} \mathbf{Q}_G^{-1} \mathbf{q}_{Gu} \overset{a}{\sim} MN\left\{\mathbf{0}, \begin{bmatrix} Q_Z^{11} & 0 & \mathbf{0} & \mathbf{0} \\ 0 & 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_W^{-1} \end{bmatrix}\right\},$$

where Q_Z^{11} is the (1,1) element of $\mathbf{Q}_Z^{-1}$. The above result can be rewritten separately for $\hat{\alpha}_{0T}$, $\hat{\mathbf{c}}_T$ and $\hat{\boldsymbol{\beta}}_T^*$ as

$$\sqrt{T}\left(\hat{\alpha}_{0T}-\alpha_0\right)+\left(\mu_1,\mu_2,...,\mu_p\right)\sqrt{T}\left(\hat{\boldsymbol{\phi}}_T-\boldsymbol{\phi}\right)=d_{Zu,1}+o_p(1), \tag{A.14}$$

$$\sqrt{T}\left(\hat{\mathbf{c}}_T-\mathbf{c}\right)+\boldsymbol{\lambda}\boldsymbol{\tau}_p'\sqrt{T}\left(\hat{\boldsymbol{\phi}}_T-\boldsymbol{\phi}\right)=o_p(1), \tag{A.15}$$

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}}_T^*-\boldsymbol{\beta}^*\right)+\left(\mathbf{g}_1,\mathbf{g}_2,...,\mathbf{g}_p\right)\sqrt{T}\left(\hat{\boldsymbol{\phi}}_T-\boldsymbol{\phi}\right)=\mathbf{Q}_W^{-1}\mathbf{q}_{Wu}+o_p(1), \tag{A.16}$$

where $\boldsymbol{\tau}_p$ is a $p\times 1$ vector of unity and $d_{Zu,1}$ is the first element of $\mathbf{Q}_Z^{-1}\mathbf{q}_{Zu}$. Using (3.10) in (A.15) we obtain (3.11). It is also clear from above results that the OLS estimators of α_0 and $\boldsymbol{\beta}^*$ (standardized by $\sqrt{T}$) have the (mixture) normal distributions asymptotically.

Finally, using (3.10), (3.11), and (A.13)-(A.16), it is easily seen that a consistent estimator of the variance of $\hat{\mathbf{h}}_T$ is given by $\hat{V}(\hat{\mathbf{h}}_T)=\hat{\sigma}_{uT}^2(\mathbf{P}_{G_T}'\mathbf{P}_{G_T})^{-1}$ with the rank of $\hat{V}(\hat{\mathbf{h}}_T)$ being equal to $kq+2$. ■

Proof of Theorem 3.2.

Partition $\mathbf{d}_T=(a_T,\mathbf{s}_T',\mathbf{w}_T')'$ conformably to $\mathbf{G}_T=(\boldsymbol{\tau}_T,\mathbf{S}_T,\mathbf{W}_T)$, then $\mathbf{s}_T$ is given by

$$\mathbf{s}_T=\sqrt{T}\left(\hat{\mathbf{c}}_T-\mathbf{c}\right)+\boldsymbol{\lambda}\boldsymbol{\tau}_p'\sqrt{T}\left(\hat{\boldsymbol{\phi}}_T-\boldsymbol{\phi}\right). \tag{A.17}$$

Using (A.10) and (A.11), $(\mathbf{s}_T',\mathbf{w}_T')'$ can be expressed as

$$\begin{aligned}\begin{bmatrix}\mathbf{s}_T\\ \mathbf{w}_T\end{bmatrix} &= \begin{bmatrix}\mathbf{S}_T'\mathbf{H}_T\mathbf{S}_T & \mathbf{S}_T'\mathbf{H}_T\mathbf{W}_T\\ \mathbf{W}_T'\mathbf{H}_T\mathbf{S}_T & \mathbf{W}_T'\mathbf{H}_T\mathbf{W}_T\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{S}_T'\mathbf{H}_T\mathbf{u}_T\\ \mathbf{W}_T'\mathbf{H}_T\mathbf{u}_T\end{bmatrix} \\ &\quad -\begin{bmatrix}\mathbf{S}_T'\mathbf{H}_T\mathbf{S}_T & \mathbf{S}_T'\mathbf{H}_T\mathbf{W}_T\\ \mathbf{W}_T'\mathbf{H}_T\mathbf{S}_T & \mathbf{W}_T'\mathbf{H}_T\mathbf{W}_T\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{S}_T'\mathbf{H}_T\mathbf{K}_T\\ \mathbf{W}_T'\mathbf{H}_T\mathbf{K}_T\end{bmatrix}\left(\hat{\boldsymbol{\phi}}_T-\boldsymbol{\phi}\right).\end{aligned} \tag{A.18}$$

Let

$$\mathbf{q}_{\tilde{S}_Tu_T}=\mathbf{D}_{S_T}\mathbf{S}_T'\mathbf{H}_T\mathbf{u}_T,\ \mathbf{Q}_{\tilde{S}_T}=\mathbf{D}_{S_T}\mathbf{S}_T'\mathbf{H}_T\mathbf{S}_T\mathbf{D}_{S_T},$$

where $\mathbf{D}_{S_T}=Diag(T^{-\frac{3}{2}},T^{-1}\mathbf{I}_k)$. Then, it is also easily seen that as $T\to\infty$,

$$\mathbf{q}_{\tilde{S}_Tu_T}\Rightarrow\mathbf{q}_{\tilde{S}u}=\begin{bmatrix}\int_0^1(r-\frac{1}{2})dB_u(r)\\ \int_0^1\tilde{\mathbf{B}}_e'(r)dB_u(r)\end{bmatrix}, \tag{A.19}$$

$$\mathbf{Q}_{\tilde{S}_T}\Rightarrow\mathbf{Q}_{\tilde{S}}=\begin{bmatrix}\frac{1}{12} & \int_0^1(r-\frac{1}{2})\tilde{\mathbf{B}}_e(r)dr\\ \int_0^1(r-\frac{1}{2})\tilde{\mathbf{B}}_e'(r)dr & \int_0^1\tilde{\mathbf{B}}_e'(r)\tilde{\mathbf{B}}_e(r)dr\end{bmatrix}, \tag{A.20}$$

where $\tilde{\mathbf{B}}_e(r) = \mathbf{B}_e(r) - \int_0^1 \mathbf{B}_e(r)dr$ is a k-dimensional demeaned Brownian motion on $[0,1]$. Since $\tilde{\mathbf{B}}_e(r)$ is also distributed independently of $B_u(r)$, we obtain as in (A.4),

$$\mathbf{q}_{\tilde{S}u} \overset{a}{\sim} MN\left(\mathbf{0}, \sigma_u^2 \mathbf{Q}_{\tilde{S}}\right). \tag{A.21}$$

Multiplying (A.18) by the diagonal matrix, $Diag(\mathbf{D}_{S_T}^{-1}, T^{\frac{1}{2}})$, using (A.19)-(A.21) and noting that

$$\begin{aligned} \mathbf{D}_{S_T}\mathbf{S}_T'\mathbf{H}_T\mathbf{W}_T &= O_p(1), \quad T^{-1}\mathbf{W}_T'\mathbf{H}_T\mathbf{W}_T = O_p(1), \\ \mathbf{D}_{S_T}\mathbf{S}_T'\mathbf{H}_T\mathbf{K}_T &= O_p(1), \quad T^{-\frac{1}{2}}\mathbf{W}_T'\mathbf{H}_T\mathbf{K}_T = O_p(1), \end{aligned}$$

we obtain

$$\mathbf{D}_{S_T}^{-1}\mathbf{s}_T \Rightarrow \mathbf{Q}_{\tilde{S}}^{-1}\mathbf{q}_{\tilde{S}u} \overset{a}{\sim} MN\left(\mathbf{0}, \sigma_u^2 \mathbf{Q}_{\tilde{S}}^{-1}\right),$$

and therefore,

$$\mathbf{Q}_{\tilde{S}_T}^{\frac{1}{2}}\mathbf{D}_{S_T}^{-1}\mathbf{s}_T \overset{a}{\sim} N\left(\mathbf{0}, \sigma_u^2 \mathbf{I}_{k+1}\right). \tag{A.22}$$

Finally, by (3.13) and (A.15) we have

$$\hat{\boldsymbol{\lambda}}_T - \boldsymbol{\lambda} = \frac{\mathbf{s}_T}{\hat{\phi}_T(1)}. \tag{A.23}$$

Multiplying (A.23) by $\mathbf{Q}_{\tilde{S}_T}^{\frac{1}{2}}\mathbf{D}_{S_T}^{-1}$, using (A.22) and noting that $\hat{\phi}_T(1) \xrightarrow{p} \phi(1)$, we obtain (3.14). ■

Proof of Theorem 3.3 can be established in a similar manner and is omitted to save space.

Proof of Theorem 4.1.

Consider the dynamic ARDL(p,m) model (3.19) (or (4.4)), and its static counterpart (4.1). Applying the decomposition $\phi(L) = \phi(1) + (1-L)\phi^*(L)$ to (3.19) we have

$$y_t = \frac{\alpha_0}{\phi(1)} + \delta t + \boldsymbol{\theta}'\mathbf{x}_t + \frac{\boldsymbol{\pi}'(L)}{\phi(1)}\Delta\mathbf{x}_t + \frac{\eta_t}{\phi(1)} - \frac{\phi^*(L)}{\phi(1)}\Delta y_t. \tag{A.24}$$

Substituting for $\Delta y_t = \delta + \boldsymbol{\theta}'\Delta\mathbf{x}_t + \Delta v_t$ from (4.1) in (A.24), we have

$$y_t = \mu + \delta t + \boldsymbol{\theta}'\mathbf{x}_t + \frac{\boldsymbol{\pi}'(L)}{\phi(1)}\Delta\mathbf{x}_t + \frac{\eta_t}{\phi(1)} - \frac{\phi^*(L)}{\phi(1)}\left(\boldsymbol{\theta}'\Delta\mathbf{x}_t + \Delta v_t\right). \tag{A.25}$$

Using (A.25), v_t in (4.1) can be expressed as

$$v_t = \frac{\boldsymbol{\pi}'(L) - \phi^*(L)\boldsymbol{\theta}'}{\phi(1)}\Delta\mathbf{x}_t + \frac{\eta_t}{\phi(1)} - \frac{\phi^*(L)}{\phi(1)}\Delta v_t. \tag{A.26}$$

Defining $\mathbf{k}_t = (\eta_t, v_t, \Delta\mathbf{x}_t')' = (\eta_t, v_t, \mathbf{e}_t')'$, and $\mathbf{\Psi}(L) = \left[\frac{1}{\phi(1)}, \frac{-\phi^*(L)(1-L)}{\phi(1)}, \frac{\boldsymbol{\pi}'(L)-\phi^*(L)\boldsymbol{\theta}'}{\phi(1)}\right]$, then the spectral density of $v_t = \mathbf{\Psi}(L)\mathbf{k}_t$ is given by

$$2\pi f_{vv}(\omega) = \mathbf{\Psi}(e^{iw})Var(\mathbf{k}_t)\mathbf{\Psi}'(e^{-iw}),$$

where

$$Var(\mathbf{k}_t) = \begin{bmatrix} \sigma_\eta^2 & \sigma_{\eta v} & \mathbf{0} \\ \sigma_{\eta v}' & \sigma_v^2 & \mathbf{\Sigma}_{ve} \\ \mathbf{0} & \mathbf{\Sigma}_{ve}' & \mathbf{\Sigma}_{ee} \end{bmatrix}.$$

Hence, the spectral density of v_t at zero frequency is given by

$$2\pi f_{vv}(0) = \frac{\sigma_\eta^2 + [\boldsymbol{\pi}'(1) - \phi^*(1)\boldsymbol{\theta}']\,\mathbf{\Sigma}_{ee}\,[\boldsymbol{\pi}(1) - \phi^*(1)\boldsymbol{\theta}]}{[\phi(1)]^2}. \qquad \text{(A.27)}$$

The Phillips-Hansen semi-parametric correction is equivalent to removing the second part of (A.27), by subtracting the terms involving $\Delta\mathbf{x}_t$ from v_t. Using (A.26) we have the following expression for the modified disturbance term, v_t^+, in the Phillips-Hansen's procedure:

$$v_t^+ = v_t - \frac{\boldsymbol{\pi}'(L) - \phi^*(L)\boldsymbol{\theta}'}{\phi(1)}\Delta\mathbf{x}_t = \frac{\eta_t}{\phi(1)} - \frac{\phi^*(L)}{\phi(1)}\Delta v_t = \mathbf{\Psi}^+(L)\mathbf{k}_t^+,$$

where $\mathbf{k}_t^+ = (\eta_t, v_t)'$, and $\mathbf{\Psi}^+(L) = \left[\frac{1}{\phi(1)}, \frac{-\phi^*(L)(1-L)}{\phi(1)}\right]$. Therefore, the spectral density of v_t^+ at zero frequency is given by

$$2\pi f_{v^+v^+}(0) = \mathbf{\Psi}^+(0)Var(\mathbf{k}_t^+)\mathbf{\Psi}^{+'}(0) = \frac{\sigma_\eta^2}{[\phi(1)]^2}.$$

Using (4.7) we also have

$$f_{v^+v^+}(0) = \mathbf{B}f_{\xi\xi}(0)\mathbf{B}',$$

where $\mathbf{B} = [1, -\mathbf{\Omega}_{ve}\mathbf{\Omega}_{ee}^{-1}]$. By definition $\mathbf{\Omega}_\xi = 2\pi f_{\xi\xi}(0)$, and

$$2\pi f_{v^+v^+}(0) = \mathbf{B}\mathbf{\Omega}_\xi\mathbf{B}' = \omega_{vv} - \mathbf{\Omega}_{ve}\mathbf{\Omega}_{ee}^{-1}\mathbf{\Omega}_{ev} = \frac{\sigma_\eta^2}{[\phi(1)]^2}.$$

Hence, by (4.8) $\omega_{v\cdot e} = \sigma_\eta^2 / [\phi(1)]^2$. ■

References

[1] Banerjee, A., J. Dolado, D. Hendry and G. Smith (1986), "Exploring Equilibrium Relationships in Economics through Statistical Models: Some Monte Carlo Evidence," *Oxford Bulletin of Economics and Statistics*, 48: 253-277.

[2] Bardsen, G. (1989), "The Estimation of Long-Run Coefficients from Error Correction Models," *Oxford Bulletin of Economics and Statistics*, 51: 345-350.

[3] Bewley, R. (1979), "The Direct Estimation of the Equilibrium Response in a Linear Dynamic Model," *Economics Letters,* 3: 357-361.

[4] Boswijk, H.P. (1995), "Efficient Inference on Cointegration Parameters in Structural Error Correction Models," *Journal of Econometrics,* 69: 133-158.

[5] Cavanaugh, C.L., G. Elliott and J.H. Stock (1995), "Inference in Models with Nearly Integrated Regressors," *Econometric Theory*, 11: 1131-1147.

[6] Engle, R.F. and C.W.J. Granger (1987), "Cointegration and Error Correction Representation: Estimation and Testing," *Econometrica*, 55: 251-276.

[7] Hendry, D., A. Pagan and J. Sargan (1984), "Dynamic Specifications'" Chapter 18 in *Handbook of Econometrics*, Vol II (ed., Z. Griliches and M. Intriligator), North Holland

[8] Hsiao, C. (1995), "Cointegration and Dynamic Simultaneous Equations Model," unpublished manuscript, University of Southern California.

[9] Inder, B. (1993), "Estimating Long Run Relationships in Economics," *Journal of Econometrics,* 57: 53-68.

[10] Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59: 1551-80.

[11] Lütkepohl H. (1991), *Introduction to Multiple Time Series Analysis*, New York, N.Y. Springer and Verlag.

[12] Park, J.Y. (1992), "Canonical Cointegrating Regressions," *Econometrica,* 60: 119-143.

[13] Pesaran, M.H. (1997), "The Role of Economic Theory in Modelling the Long-Run," *The Economic Journal,* 107: 178-191.

[14] Pesaran, M.H. and B. Pesaran (1997), *Microfit 4.0: Interactive Econometric Analysis*, Oxford University Press (forthcoming).

[15] Pesaran, M.H. and Y. Shin (1995), "Long-Run Structural Modelling," unpublished manuscript, University of Cambridge.

[16] Pesaran, M.H., Y. Shin and R.J. Smith (1996), "Testing for the Existence of a Long-Run Relationship," *DAE Working Papers Amalgamated Series*, No. 9622, University of Cambridge.

[17] Phillips, P.C.B. (1991), "Optimal Inference in Cointegrated Systems," *Econometrica,* 59: 283-306.

[18] Phillips, P.C.B. and S.N. Durlauf (1986), "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies,* 53: 473-496.

[19] Phillips, P.C.B. and B. Hansen (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies,* 57: 99-125.

[20] Phillips, P.C.B. and M. Loretan (1991), "Estimating Long Run Economic Equilibria," *Review of Economic Studies,* 58: 407-436.

[21] Phillips, P.C.B. and V. Solo (1992), "Asymptotic for Linear Processes," *Annals of Statistics*: 971-1001.

[22] Saikkonen, P (1991), "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory,* 7: 1-21.

[23] Stock, J.H. (1987), "Asymptotic Properties of Least Squares Estimates of Cointegrating Vectors," *Econometrica*, 55: 1035-1056.

[24] Stock, J.H. and M.W. Watson (1993), "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61: 783-820.

[25] Wickens, M.R. and T.S. Breusch (1988), "Dynamic Specification, the Long Run Estimation of the Transformed Regression models," *The Economic Journal,* 98: 189-205.