Allgemeines Statistisches Archiv 0, 0–15
 ISSN 0002-6018

Autoregressive Distributed Lag Models and Cointegration

By Uwe Hassler and Jürgen Wolters*

Summary: This paper considers cointegration analysis within an autoregressive distributed lag (ADL) framework. First, different reparameterizations and interpretations are reviewed. Then we show that the estimation of a cointegrating vector from an ADL specification is equivalent to that from an error-correction (EC) model. Therefore, asymptotic normality available in the ADL model under exogeneity carries over to the EC estimator. Next, we review cointegration tests based on EC regressions. Special attention is paid to the effect of linear time trends in case of regressions without detrending. Finally, the relevance of our asymptotic results in finite samples is investigated by means of computer experiments. In particular, it turns out that the conditional EC model is superior to the unconditional one.

Keywords: Error-correction, asymptotically normal inference, cointegration testing. JEL C22, C32.

1. Introduction

The autoregressive distributed lag model (ADL) is the major workhorse in dynamic single-equation regressions. One particularly attractive reparameterization is the error-correction model (EC). Its popularity in applied time series econometrics has even increased, since it turned out for nonstationary variables that cointegration is equivalent to an error-correction mechanism, see Granger's representation theorem in Engle and Granger (1987). By differencing and forming a linear combination of the nonstationary data, all variables are transformed equivalently into an EC model with stationary series only.

Working on feedback control mechanisms for stabilization policy, Phillips (1954, 1957) introduced EC models to economics. Sargan (1964) used them to estimate structural equations with autocorrelated residuals, and Hendry popularized their use in econometrics in a series of papers[1]. According to Hylleberg and Mizon (1989, p.124) "*the error correction formulation provides an excellent framework within which it is possible to apply both the data information and the information available from economic theory*". A survey on specification, estimation and testing of EC models is given by Alogoskoufis and Smith (1995). The present paper contributes to this literature in that it treats some aspects of testing cointegration and asymptotic normal inference of the cointegrating vector estimated from an EC format.

Received: / Revised:

* We thank Vladimir Kuzin for excellent research assistance and Surayyo Kabilova for skillful word processing. Moreover, we are grateful to an anonymous referee for clarifying comments.

[1] Davidson et al. (1978), Hendry (1979), and Hendry et al. (1984). It is noteworthy that A.W. Phillips, Sargan as well as Hendry were professors at the London School of Economics. A personal view on the history of EC models is given in the interview of Hendry by Ericsson (2004).

The rest of the paper is organized as follows. The next section reviews different reparameterizations and interpretations of ADL models. Then we use that the cointegrating vector computed from the ADL model is equivalent to the one estimated from EC in order to use results by Pesaran and Shin (1998) on asymptotic normality. Section 4 turns to cointegration testing from EC regressions. We review t-type and F-type test statistics, and pay particular attention to the role of linear time trends. The relevance of our asymptotic results in finite samples is investigated through Monte Carlo experiments in Section 5. A detailled summary is contained in the final section.

2. Assumptions and representations

The autoregressive distributed lag model of order p and n, ADL(p,n), is defined for a scalar variable y_t as

$$y_t = \sum_{i=1}^{p} a_i\, y_{t-i} + \sum_{i=0}^{n} c_i'\, x_{t-i} + \varepsilon_t, \tag{1}$$

where ε_t is a scalar zero mean error term and x_t is a K-dimensional column vector process. Typically, a constant is included in (1), which we neglect here for brevity. The coefficients a_i are scalars while c_i' are row vectors. Using the lag operator L applied to each component of a vector, $L^k\, x_t = x_{t-k}$, it is convenient to define the lag polynomial $a(L)$ and the vector polynomial $c(L)$,

$$\begin{aligned} a(L) &= 1 - a_1 L - \ldots - a_p\, L^p, \\ c(L) &= c_0 + c_1 L + \ldots + c_n\, L^n. \end{aligned}$$

Now, it is straightforward to write (1) more compactly:

$$a(L) y_t = c'(L) x_t + \varepsilon_t.$$

In order to obtain dynamic stability, it is maintained that

$$a(z) = 0 \quad \Rightarrow \quad |\, z \,| > 1 \;\text{ for } z \in \mathbb{C}. \tag{2}$$

Under this condition there exists an absolutely summable infinite expansion of the inverted polynomial $a^{-1}(L)$:

$$a^{-1}(L) = \frac{1}{a(L)} = \sum_{j=0}^{\infty} a_j^*\, L^j, \quad \sum_{j=0}^{\infty} |a_j^*| < \infty.$$

Invertibility of $a(L)$ hence yields the following representation:

$$y_t = \frac{c'(L)}{a(L)}\, x_t + e_t, \quad a(L)\, e_t = \varepsilon_t,$$

where e_t has a stable autoregressive structure of order p. Expanding $a^{-1}(L)$ provides an infinite distributed lag representation,

$$y_t = \left(\sum_{j=0}^{\infty} a_j^* L^j \right) \left(\sum_{j=0}^{n} c_j \, L^j \right)' x_t + e_t = \sum_{j=0}^{\infty} b_j' \, x_{t-j} + e_t, \tag{3}$$

where b_j are the vectors of dynamic multipliers derived by the method of indetermined coefficients. The vector of long-run multipliers of the ADL(p,n) model may therefore be easily computed from:

$$\beta := \frac{c(1)}{a(1)} = \sum_{j=0}^{\infty} b_j . \tag{4}$$

It is worth mentioning that (1) is suitable for estimation but in order to obtain an economic interpretation of the parameters one has to consider a transformation like (3).

Different reparameterizations have been discussed in the literature, see e.g. Wickens and Breusch (1988). By re-arranging the x's one obtains with $\Delta = 1 - L$:

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + a(1)\beta' x_t - \sum_{i=0}^{n-1} \left(\sum_{j=i+1}^{n} c_j \right)' \Delta x_{t-i} + \varepsilon_t, \tag{5}$$

where y_t is related to its own past, to contemporaneous x_t and differences Δx_{t-i}. The use of this specification has been suggested for cointegration analysis by Pesaran and Shin (1998). A further variant relates y_t to x_t and differences of both variables. By subtracting $(\sum_{i=1}^{p} a_i)\, y_t$ and re-normalizing, (5) yields:

$$y_t = \frac{-1}{a(1)} \sum_{i=0}^{p-1} \left(\sum_{j=i+1}^{p} a_j \right) \Delta y_{t-i} + \beta' x_t - \frac{1}{a(1)} \sum_{i=0}^{n-1} \left(\sum_{j=i+1}^{n} c_j \right)' \Delta x_{t-i} + \varepsilon_t .$$

This representation due to Bewley (1979) has the advantage that the long-run multipliers β are the coefficients of x_t. However, the contemporaneous Δy_t on the right-hand side is correlated with ε_t, which renders OLS invalid. Nevertheless, the use of $y_{t-1}, \ldots, y_{t-p-1}$ and $x_t, \ldots, x_{t-n+1}$ as instruments allows for consistent instrumental variable estimation.

One further transformation will turn out to be fruitful for cointegration testing and estimation. Notice that

$$\sum_{i=1}^{p} a_i y_{t-i} - y_{t-1} = -a(1) y_{t-1} - \sum_{i=1}^{p-1} \left(\sum_{j=i+1}^{p} a_j \right) \Delta y_{t-i} \, .$$

Using this result and $x_t = x_{t-1} + \Delta x_t$, (5) yields the error-correction format:

$$\Delta y_t = -a(1)\,(y_{t-1} - \beta' x_{t-1}) - \sum_{i=1}^{p-1} \left(\sum_{j=i+1}^{p} a_j \right) \Delta y_{t-i}$$
$$+ \left(a(1)\beta - \sum_{j=1}^{n} c_j \right)' \Delta x_t - \sum_{i=1}^{n-1} \left(\sum_{j=i+1}^{n} c_j \right)' \Delta x_{t-i} + \varepsilon_t .$$

The interpretation relies on a long-run equilibrium relation, $y = \beta' x$. The error-correction mechanism is the adjustment of y_t via $a(1)$ to equilibrium deviations in the previous period, $y_{t-1} - \beta' x_{t-1}$. In the following, this equation will often be rewritten as

$$\Delta y_t = \gamma\, y_{t-1} + \theta' x_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta y_{t-i} + \sum_{i=0}^{n-1} \phi_i' \Delta x_{t-i} + \varepsilon_t, \tag{6}$$

where

$$\gamma = -a(1), \quad \theta = a(1)\,\beta = -\gamma\,\beta, \tag{7}$$

and α_i as well as ϕ_i are defined in an obvious manner.

Since the work by Engle and Granger (1987), cointegration of nonstationary processes is known to be equivalent to a data generating error-correction process. For the rest of the paper we assume that y_t and x_t are integrated of order one, I(1), i.e. differencing is required to obtain stationarity. When there exists a linear combination of the nonstationary processes, $y_t - \beta' x_t$, $\beta \neq 0$, which is stationary, then y_t and x_t are called cointegrated. The cointegration rank is at most one, and x_t does not adjust towards equilibrium.

Assumption 1: *(i) The vector $(y_t, x_t')'$ of length $K+1$ is $I(1)$. (ii) The vector x_t alone is not cointegrated. (iii) In case of cointegration, x_t does not adjust to past equilibrium deviations $(y_{t-1} - \beta'\, x_{t-1})$.*

Further, we assume a correctly specified error-correction equation in the following sense.

Assumption 2: *(i) The errors ε_t are serially independent with variance σ^2, $\varepsilon_t \sim iid(0, \sigma^2)$. (ii) The errors are uncorrelated with $\Delta\, x_{t+h}$, for all $h \in \mathbb{Z}$.*

These assumptions summarize (A1) through (A5) in Pesaran and Shin (1998, p.375). The case of several linearly independent cointegrating vectors or the situation where Δx_t adjusts to lagged deviations, too, is beyond the scope of a single-equation framework, see e.g. Lütkepohl (2006) in this volume.

Assumption 2 (ii) was made to ensure exogeneity of Δx_t. It may seem very restrictive for applied work. Working with normally distributed data, however, we do not need it because Johansen (1992) proved assuming a Gaussian vector EC model for $(y_t, x_t')'$ that Assumption 1 (iii) alone is sufficient for weak exogeneity of Δx_t, cf. also Urbain (1992. Prop.1). In fact, he thus showed that under Assumption 1 (iii) alone the single-equation analysis is equivalent to maximum likelihood estimation of the full system (Johansen, 1992, Corollary 1).

3. Inference about the cointegrating vector

In this section we assume that y_t and x_t are cointegrated, and the interest focusses on estimating and testing β given T observations. It is well known since Phillips and Durlauf (1986) or Stock (1987) that the static OLS estimator,

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}'\, x_t, \quad t = 1, \dots, T,$$

is super-consistent. Under exogeneity, it further holds (cf. Phillips and Park, 1988) that $T(\hat{\beta} - \beta)$ converges to a normal distribution, where the variance depends on the long-run variance (or spectral density at frequency zero) of $y_t - \beta'\, x_t$. This parameter may be difficult to estimate in finite samples. Moreover, already Banerjee et al. (1986) observed that static OLS may be biased in finite samples due to ignoring short-run dynamics. An alternative approach dating back to Stock (1987) relies on estimating (6):

$$\Delta y_t = \hat{c} + \hat{\gamma}\, y_{t-1} + \hat{\theta}'\, x_{t-1} + \sum_{i=1}^{p-1} \hat{\alpha}_i\, \Delta y_{t-i} + \sum_{i=0}^{n-1} \hat{\phi}_i'\, \Delta x_{t-i} + \hat{\varepsilon}_t\,. \tag{8}$$

A natural candidate for estimating β is now from (8) because of (7)

$$\hat{\beta}_{EC} = -\frac{\hat{\theta}}{\hat{\gamma}}. \tag{9}$$

Further down we will obtain limiting normality of $T(\hat{\beta}_{EC} - \beta)$ under exogeneity by drawing upon results by Pesaran and Shin (1998), who consider the OLS estimation of (5):

$$y_t = \tilde{c} + \sum_{i=1}^{p} \tilde{a}_i\, y_{t-i} + \tilde{\theta}' x_t + \sum_{i=0}^{n-1} \tilde{\phi}_i'\, \Delta x_{t-i} + \tilde{\varepsilon}_t. \tag{10}$$

As estimator for β they propose because of (7):

$$\hat{\beta}_{PS} = \frac{\tilde{\theta}}{1 - \sum_{i=1}^{p} \tilde{a}_i}\,. \tag{11}$$

Pesaran and Shin (1998, Theorem 2.4 or 3.2) establish limiting normality under the stated assumptions.

Proposition 1: *Under Assumptions 1 and 2 and under cointegration it holds as $T \to \infty$:*

$$\left[\sum_{t=1}^{T} (x_t - \bar{x})\,(x_t - \bar{x})' \right]^{0.5} \left(\hat{\beta}_{PS} - \beta \right) \;\sim\; \mathcal{N}_K \left(0,\, \frac{\sigma^2}{(a(1))^2} I_K \right),$$

where I_K denotes the identity matrix.

REMARK A It is noteworthy that $\sum (x_t - \bar{x})(x_t - \bar{x})'$ diverges with T^2, so that $\hat{\beta}_{PS}$ converges with the expected super-consistent rate T. Although normality arises just like in the stationary case, the rate of convergence differs from the situation where x_t is I(0). Moreover, σ^2 and $a(1)$ may be estimated consistently:

$$\tilde{\sigma}^2 = \frac{1}{T - m} \sum_{t=1}^{T} \tilde{\varepsilon}_t^2, \quad \tilde{a}(1) = 1 - \sum_{i=1}^{p} \tilde{a}_i\,,$$

where $m = K(n+1) + p + 1$ denotes the number of estimated parameters including a constant. Finally, by demeaning x_t in Proposition 1, we assume that the regression equation contains an intercept. The result continues to hold, if a linear time trend as additional regressor is allowed for.

REMARK B In practice, Assumption 2 (ii) may be too restrictive, and (lagged values of) Δx_t may be correlated with ε_t. To account for that, Pesaran and Shin (1998) propose to simply include the corresponding difference Δx_{t-k} as additional regressor in (10) in case that $k \geq n$.

Since (6) is a linear transformation of (5), it turns out that the regression (8) is a linear transformation of (10). Using the techniques by Wickens and Breusch (1988) we can establish the following result. The proof is tedious but not difficult, details are available upon request.

Proposition 2: *For the OLS regressions (10) and (8) it holds:*

$$\hat{\gamma} = \sum_{i=1}^{p} \tilde{a}_i - 1, \quad \hat{\theta} = \tilde{\theta}, \;\; \hat{\varepsilon}_t = \tilde{\varepsilon}_t,$$

and consequently: $\hat{\beta}_{EC} = \hat{\beta}_{PS}$.

As a corollary to Propositions 1 and 2, $\hat{\beta}_{EC}$ follows a limiting normal distribution. Consider a t type statistic testing for the kth component $\beta^{(k)}$, $k = 1, 2, \ldots, K$:

$$\tau_k = \frac{|\hat{\gamma}|\,(\hat{\beta}_{EC}^{(k)} - \beta^{(k)})}{\hat{\sigma} \sqrt{[(\sum (x_t - \bar{x})(x_t - \bar{x})')^{-1}]_{kk}}},$$

where $[\cdot]_{kk}$ denotes the entry on the principal diagonal of a matrix, and, obviously: $\hat{\sigma}^2 = \frac{1}{T-m}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$ where again $m = K(n+1)+p+1$.

Corollary 1: *Under the assumptions of Proposition 1 it holds for $k = 1, \ldots, K$:*

$$\tau_k \ \sim \ \mathcal{N}(0,1),$$

as $T \to \infty$.

Concluding this section it should be noticed that estimation of and inference about β from linear or nonlinear dynamic regressions similar to (8) and (10) has been discussed by Stock (1987), Phillips (1988), Phillips and Loretan (1991), and Boswijk (1995), too.

4. Cointegration testing

We consider tests for the null hypothesis of no cointegration building on the error-correction equation (6) augmented by a constant intercept and estimated by OLS, $t = 1, 2, \ldots, T$,

$$\Delta y_t = \widehat{c} + \widehat{\gamma}_\mu\, y_{t-1} + \widehat{\theta}'_\mu x_{t-1} + \sum_{i=1}^{p-1}\widehat{\alpha}_i \Delta y_{t-i} + \sum_{i=0}^{n-1}\widehat{\phi}'_i \Delta x_{t-i} + \widehat{\varepsilon}_t. \quad (12)$$

Sometimes empirical researchers wish to work with detrended series, which amounts to adding a linear time trend to the set of regressors[2]:

$$\Delta y_t = \widehat{c} + \widehat{\delta}\, t + \widehat{\gamma}_\tau\, y_{t-1} + \widehat{\theta}'_\tau x_{t-1} + \sum_{i=1}^{p-1}\widehat{\alpha}_i \Delta y_{t-i} + \sum_{i=0}^{n-1}\widehat{\phi}'_i \Delta x_{t-i} + \widehat{\varepsilon}_t. \quad (13)$$

Clearly, the linear trend will change all parameter estimates. For that reason γ and θ are now indexed with τ, while all other estimates are denoted by the same symbols as in (12) for convenience. Sometimes, (12) and (13) are called conditional (or structural) error-correction models, while unconditional (reduced form) models are obtained by restricting $\phi_0 = 0$ and excluding contemporaneous differences, Δx_t.

Given Assumption 1, the null hypothesis of no cointegration[3] may be parameterized as follows:

$$H_0 : \gamma_\mu = 0 \quad \text{or} \quad \gamma_\tau = 0.$$

[2] This is not so much a question of choosing the "right model" but rather the economically more meaningful one, cf. Hassler (1999) for a discussion.

[3] Pesaran et al. (2001) consider a more general procedure where the order of integration of x_t is not assumed to be known. The null hypothesis then reads as "there is no stable level relationship between y_t and x_t". The limiting distribution depends on the I(0) or I(1) assumption, but Pesaran et al. (2001) propose to use critical values from both distributions as bounds: If the test statistic falls outside the bounds, a conclusive inference may be drawn without having to know the integration status of the underlying variables.

Under the alternative of cointegration equilibrium adjustment implies

$$H_1 : \gamma_\mu < 0 \quad \text{or} \quad \gamma_\tau < 0.$$

Therefore, Banerjee et al. (1998) proposed the use of the conventional studentized t statistic relying on an OLS estimation of (12) or (13):

$$ECt_\mu = t_{\gamma_\mu = 0} \quad \text{or} \quad ECt_\tau = t_{\gamma_\tau = 0}.$$

The null hypothesis is rejected for too small (negative) values.

Similarly, Boswijk (1994) suggested an F type test for

$$H_0 : \gamma_\mu = 0,\, \theta_\mu = 0 \quad \text{or} \quad \gamma_\tau = 0,\, \theta_\tau = 0.$$

Let $F_{\gamma,\theta}$ denote the conventional F statistics from (12) or (13) testing for lack of significance. Then Boswijk (1994) considered

$$ECF_\mu = (K+1) F_{\gamma_\mu, \theta_\mu} \quad \text{or} \quad ECF_\tau = (K+1) F_{\gamma_\tau, \theta_\tau}.$$

Here, the null hypothesis is rejected for too large values. Boswijk (1994) suggested a further variant for (13), where the linear trend is restricted under H_0:

$$H_0 : \gamma_\tau = 0, \quad \theta_\tau = 0, \quad \delta = 0.$$

The corresponding F type statistic tests for $K+2$ restrictions:

$$ECF_\tau^* = (K+2) F_{\gamma_\tau, \theta_\tau, \delta}.$$

In many economic applications it may occur that x_t is $I(1)$ with drift,

$$E(\Delta x_t) = d \neq 0.$$

Still, empirical workers often wish to regress without detrending. However, the linear trend is growing with T while the stochastic trend is only of order $T^{0.5}$. Hence, the linear in the data,

$$\begin{aligned} x_t &= x_0 + d\, t + I(1) \\ &= O_p(1) + O_p\,(T) + O_p\,(T^{0.5}), \end{aligned}$$

dominates the stochastic trend and hence affects the limiting distribution of ECt_μ from (12). Fortunately, critical values are nevertheless readily available.

Proposition 3: *Under Assumptions 1 and 2 and the null hypothesis of no cointegration, it holds as $T \to \infty$:*

a) $ECt_\tau \xrightarrow{d} \mathcal{BDM}_\tau(K)$ *for any* $E(\Delta x_t)$;

b) $ECt_\mu \xrightarrow{d} \mathcal{BDM}_\mu(K)$ *for* $E(\Delta x_t) = 0$;

c) $ECt_\mu \xrightarrow{d} \mathcal{BDM}_\tau(K-1)$ *for* $E(\Delta x_t) \neq 0$, *where* $\mathcal{BDM}_\tau(0)$ *stands for the detrended Dickey-Fuller distribution.*

Convergence in distribution is denoted by $\xrightarrow{d}$. The variables $\mathcal{BDM}_\mu(K)$ and $\mathcal{BDM}_\tau(K)$ represent functionals of vector standard Brownian motions of length K, which are demeaned and detrended, respectively. K denotes the number of variables contained in the vector x_t. Detailed expressions of those limiting distributions and simulated critical values can be found in Banerjee et al. (1998), who prove a) and b). The third result was established by Hassler (2000), and by detrended Dickey-Fuller distribution we mean the limit of $\widehat{\tau}_\tau$ in the notation by Dickey and Fuller (1979).

REMARK C In applied work it may be not so clear whether x_t is dominated by a linear trend or not. It hence may be dubious whether to use Proposition 3 b) or c) for inference. Looking at corresponding percentiles in Table 1, we learn that critical values from $\mathcal{BDM}_\tau(K-1)$ are smaller than those from $\mathcal{BDM}_\mu(K)$ for K being fixed. Usage of $\mathcal{BDM}_\tau(K-1)$ in case of regressions without detrending therefore results in a conservative test avoiding over-rejection, which may be the advisable strategy in practice.

Similar results to Proposition 3 are available for the F type statistics.

Proposition 4: *Under Assumptions 1 and 2 and the null hypothesis of no cointegration, it holds as $T \to \infty$:*

a) $ECF_\tau \xrightarrow{d} \mathcal{B}_\tau(K)$ *for any* $E(\Delta x_t)$*;*
b) $ECF_\tau^* \xrightarrow{d} \mathcal{B}_\tau^*(K)$ *for any* $E(\Delta x_t)$*;*
c) $ECF_\mu \xrightarrow{d} \mathcal{B}_\mu(K)$ *for* $E(\Delta x_t) = 0$*.*

Boswijk (1994) characterized the stochastic limits of the type $\mathcal{B}$ depending again on the number of I(1)-variables x_t and on the deterministics (with or without linear trend). However, there remains one question. How do linear trends in the data affect the limiting distribution of the F type test ECF_μ without detrending? Without proof we state motivated by Proposition 3 c) the following conjecture (a proof could follow the lines in Hassler, 2000).

Conjecture *When $E(\Delta x_t) \neq 0$, we conjecture under the assumptions of Proposition 4 for the regression without detrending:*

$$ECF_\mu \xrightarrow{d} \mathcal{B}_\tau^*(K-1), \tag{14}$$

where in case of $K = 1$, $\mathcal{B}_\tau^*(0)$ is understood to be twice the limiting distribution of the Φ_3 statistic from Dickey and Fuller (1981); see Table VI in Dickey and Fuller (1981) for percentiles: $\Phi_3 \xrightarrow{d} \frac{1}{2}B_\tau^*(0)$.

The applicability of (14) in finite samples will be established by computer experiments in Section 5. The intuition behind this claim is again that the process x_t follows one common linear time trend and K stochastic I(1) trends. The linear trend dominates one stochastic trend. Therefore, in case of linear trends it holds the following asymptotically: testing for $\theta_\mu = 0$ in (12) with θ_μ being of length K amounts to the same as if we tested for $\delta = 0$ and $\theta_\tau = 0$ where θ_τ was only $(K-1)$-dimensional in (13).

Examples of critical values of the distributions encountered in this section are given in Table 1. We observe that critical values of $\mathcal{B}_\tau^*(K-1)$ are shifted to the right relative to $\mathcal{B}_\mu(K)$. Analogously to Remark C the use of $\mathcal{B}_\tau^*(K-1)$ instead of $\mathcal{B}_\mu(K)$ in case of regressions without detrending hence results in a conservative procedure.

Table 1: Critical values

	$\mathcal{BDM}_\mu(K)$	$\mathcal{BDM}_\tau(K-1)$	$\mathcal{B}_\mu(K)$	$\mathcal{B}_\tau^*(K-1)$
	$K=1$			
1 %	-3.78	-3.96	15.22	16.54
5 %	-3.19	-3.41	11.41	12.50
10 %	-2.89	-3.13	9.54	10.68
	$K=2$			
1 %	-4.06	-4.27	18.68	19.30
5 %	-3.48	-3.69	14.38	15.24
10 %	-3.19	-3.39	12.22	13.22
	$K=3$			
1 %	-4.46	-4.51	21.43	22.50
5 %	-3.74	-3.91	17.18	18.03
10 %	-3.42	-3.62	14.93	15.85
	$K=4$			
1 %	-4.57	-4.72	24.63	25.46
5 %	-3.97	-4.12	19.69	20.66
10 %	-3.66	-3.82	17.38	18.45
	$K=5$			
1 %	-4.70	-4.89	27.11	28.51
5 %	-4.27	-4.30	22.48	23.33
10 %	-3.82	-4.00	19.87	20.76

Note: The asymptotic critical values of $\mathcal{BDM}_\mu(K)$ and $\mathcal{BDM}_\tau(K-1)$ are taken from Banerjee et al. (1998, Table I), except for $\mathcal{BDM}_\tau(0)$ from Fuller (1996, Table 10.A.2). The percentiles of $\mathcal{B}_\mu(K)$ and $\mathcal{B}_\tau^*(K-1)$ are from Tables B.2 and B.5 in Boswijk (1994), except for $\mathcal{B}_\tau^*(0)$. The latter quantiles are twice the values found in Dickey and Fuller (1981, Table VI).

5. Monte Carlo Evidence

For simulation purposes we generated a bivariate process ($K = 1$) as

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} -\gamma_1 \\ \gamma_2 \end{pmatrix} \begin{pmatrix} y_{t-1} - x_{t-1} \end{pmatrix} + \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{pmatrix} + \varepsilon_t \,, \qquad (15)$$

$$\varepsilon_t \sim ii\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad t = 1, 2, \ldots, T \,. \qquad (16)$$

We consider the conditional error-correction regression,

$$\Delta y_t = \hat{c} + \hat{\gamma}\, y_{t-1} + \hat{\theta}\, x_{t-1} + \hat{\alpha}_1\, \Delta y_{t-1} + \hat{\phi}_0 \Delta x_t + \hat{\phi}_1\, \Delta x_{t-1} + \hat{\varepsilon}_t, \qquad (17)$$

as well as the unconditional one without contemporaneous Δx_t:

$$\Delta y_t = \tilde{c} + \tilde{\gamma}\, y_{t-1} + \tilde{\theta}\, x_{t-1} + \tilde{\alpha}_1\, \Delta y_{t-1} + \tilde{\phi}_1\, \Delta x_{t-1} + \tilde{\varepsilon}_t. \qquad (18)$$

Clearly, the unconditional regression (18) is only appropriate when $\rho = 0$; when $\rho \neq 0$, however, the inclusion of Δx_t is required to account for simultaneous correlation.

Throughout we present rejections at the nominal 5 % level that are obtained from 50000 replications. All programming[4] was done in Ox Professional 3.30.

Table 2: Asymptotically normal cointegration vector

	$T = 100$			$T = 250$			$T = 1000$		
$\gamma_1 =$	0.2	0.4	0.6	0.2	0.4	0.6	0.2	0.4	0.6
$\gamma_2 = 0$	Conditional regression (17)								
$\rho = 0$	11.4	9.0	7.8	7.2	6.4	6.1	5.5	5.3	5.2
$\rho = 0.3$	11.1	9.1	8.0	7.3	6.5	6.2	5.6	5.3	5.3
$\rho = 0.6$	11.1	8.8	8.3	7.3	6.4	6.1	5.7	5.3	5.2
$\gamma_2 = 0$	Unconditional regression (18)								
$\rho = 0$	10.9	8.5	7.4	7.0	6.2	5.9	5.5	5.3	5.1
$\rho = 0.3$	12.2	10.2	9.4	9.2	8.3	8.1	7.5	7.3	7.4
$\rho = 0.6$	17.8	16.0	15.4	15.1	14.2	14.1	13.8	13.9	13.7
$\gamma_2 = \gamma_1$	Conditional regression (17)								
$\rho = 0$	22.6	20.9	21.0	19.0	18.3	18.4	17.9	17.6	17.5
$\rho = 0.3$	18.1	16.9	16.3	15.1	14.6	14.6	13.8	13.8	13.9
$\rho = 0.6$	13.7	12.7	12.5	11.4	10.7	10.5	10.2	10.0	10.1

Note: The true DGP is (15) with (16). We report the frequency of rejection of a two-sided test as in Corollary 1 at the 5 % significance level.

[4] We thank Vladimir Kuzin for computational help.

Table 2 contains results for the asymptotically normal cointegration estimator $\hat{\beta}_{EC}$, see Corollary 1. For the upper and the middle panel we assume $\gamma_2 = 0$ and $\gamma_1 \in \{0.2, 0.4, 0.6\}$. With growing γ_1 (i.e. error-correction adjustment) the experimental size improves. For $T = 100$ the test is oversized. With $T = 250$, the experimental level of the conditional regression is fairly close to the nominal one, and the correspondence is very good for $T = 1000$. Moreover, for $\rho > 0$, Assumption 2 (ii) is violated because Δx_t and the regression error are correlated. This turns the unconditional regression (18) invalid, while the conditional regression is not affected by ρ. This supports the proposal by Pesaran and Shin (1998) to add (lags of) Δx_t in case that Assumption 2 (ii) does not hold in order to maintain limiting normality, cf. Remark B. In the lower panel Assumption 1 (iii) is violated because $\gamma_2 = \gamma_1 \neq 0$. In this situation Δx_t is not exogeneous as proven by Johansen (1992). Therefore, even the conditional regression does not result in a limiting $\mathcal{N}(0,1)$ distribution as is well demonstrated for $T = 1000$.

Table 3: Cointegration tests

	$\gamma = 0$	γ_1 $(\gamma_2 = 0)$			γ_2 $(\gamma_1 = 0)$			$\gamma_1 = \gamma_2$		
	H_0	0.05	0.1	0.2	0.05	0.1	0.2	0.05	0.1	0.2
Conditional regression (17)										
	$\rho = 0$									
ECt_μ	6.1	30.0	75.4	99.6	2.7	1.5	0.7	22.1	48.1	79.8
ECF_μ	6.4	29.2	71.6	99.2	5.5	4.3	4.1	20.4	45.4	80.0
	$\rho = 0.3$									
ECt_μ	5.7	26.2	68.0	98.9	6.6	7.7	8.5	32.0	67.7	94.0
ECF_μ	6.4	23.9	64.0	98.4	8.6	10.0	13.1	30.6	66.6	94.1
	$\rho = 0.6$									
ECt_μ	5.8	25.6	67.3	98.8	14.6	26.7	45.1	48.2	87.3	99.2
ECF_μ	6.7	23.3	63.0	98.2	17.7	31.3	51.1	47.4	86.9	99.2
Unconditional regression (18)										
	$\rho = 0$									
ECt_μ	6.0	30.4	76.1	99.6	2.7	1.7	0.7	24.1	55.1	89.0
ECF_μ	6.4	29.1	72.3	99.3	5.7	4.3	4.2	22.0	51.8	88.4
	$\rho = 0.3$									
ECt_μ	5.5	22.9	60.8	96.3	2.3	1.4	0.7	16.1	35.7	70.8
ECF_μ	6.8	21.8	58.1	95.9	5.6	4.5	3.9	17.0	37.3	72.6
	$\rho = 0.6$									
ECt_μ	4.4	13.6	34.1	76.0	1.8	1.1	0.5	8.0	17.0	39.7
ECF_μ	6.4	15.3	38.7	80.9	5.5	4.5	4.1	12.4	24.4	48.4

Note: The true DGP is (15) with (16) and $T = 100$. We report the frequency of rejection at the 5 % significance level.

Table 3 displays findings for the cointegration tests ECt and ECF with $T = 100$ only. In the column "$\gamma = 0$" it holds $\gamma_1 = \gamma_2 = 0$, and the null hypothesis is true. The next three columns assume $\gamma_2 = 0$ and $\gamma_1 \in \{0.05, 0.1, 0.2\}$. The power increases with γ_1, and the t and F tests behave very similarly. The conditional regression including Δx_t produces tests that are robust with respect to ρ, while (18) results in dramatic power losses as ρ grows. In the next three columns ($\gamma_1 = 0$, $\gamma_2 \in \{0.05, 0.1, 0.2\}$) we do have cointegration but y_t does not adjust. Hence, the unconditional regression provides no power, while (17) still allows to reject, as long as $\rho \neq 0$. Only here it turns out that the F type test is slightly more powerful. In the last three columns Assumption 1 (iii) does not hold ($\gamma_2 = \gamma_1 \in \{0.05, 0.1, 0.2\}$). If $\rho \neq 0$, the power increases with growing ρ for the tests based on conditional regressions compared with $\gamma_2 = 0$, while in case of unconditional regressions the power is reduced.

Finally, Table 4 supports our Conjecture. Here, we simulated $(K+1)$-dimensional random walks $(y_t, x_t')'$ independent of each other. Moreover, x_t contains a drift, which is identical in all components:

$$x_t = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} t + \sum_{i=1}^{t} \varepsilon_i .$$

Application of ECF_μ from (17) with critical values from $\mathcal{B}_\tau^*(K-1)$ provides a valid approximation as T increases.

Table 4: Conjecture

	$T = 100$			$T = 250$			$T = 1000$		
	$K=1$	$K=2$	$K=3$	$K=1$	$K=2$	$K=3$	$K=1$	$K=2$	$K=3$
1 %	1.35	1.84	1.89	1.06	1.49	1.46	0.98	1.34	1.26
5 %	5.40	6.18	6.60	5.17	5.80	5.77	4.89	5.50	5.47
10 %	10.37	11.45	11.78	9.96	11.04	11.18	9.72	10.70	10.53

Note: The true DGP is a random walk with drift. We report rejection frequencies of the F test applied to (17) with critical values from $\mathcal{B}_\tau^*(K-1)$.

6. Summary

We reviewed different parameterizations of the autoregressive distributed lag (ADL) model and stressed the equivalence with error-correction (EC) mechanisms. This motivates the following finding: the cointegrating vector and the residuals computed from the EC model are numerically identical to the ones constructed from the ADL regression. Therefore, under the exogeneity conditions of Pesaran and Shin (1998) the limiting normality of the estimated cointegrating vector carries over to the EC model. Next, we review t-type and F-type tests for the null hypothesis of no cointegration proposed in an EC framework by Banerjee et al. (1998) and Boswijk (1994), respectively. Hassler (2000) treated the t-type test in the presence of linear trends in the data when regressions are run without detrending. Here, we treat the F-type test in the same situation. We refrain from proving the limiting distribution but support a conjecture by means of simulation evidence instead.

The main results of our Monte Carlo study are the following. First, in most cases the the t-type cointegration test is just as powerful as the F-type one. Second, we investigate the case that is of particular interest in applied work where Δx_t is correlated with the regression error. In this situation, the conditional regression (including contemporaneous Δx_t as regressor) still provides valid inference about the cointegration vector relying on the normal approximation. For this result to hold true it is crucial that Δx_t is exogenous in the sense that it does not adjust to past equilibrium deviations. Moreover, cointegration tests from the conditional regression are more powerful than those from unconditional ones. A general finding hence is that the conditional error-correction regression outperforms the unconditional one.

Bibliography

Alogoskoufis, G., Smith, R. (1995). On error correction models: Specification, interpretation, estimation. In *Surveys in Econometrics* (L. Oxley, D.A.R. George, C.J. Roberts, S. Sayer, eds.), 139-170. Blackwell Publishers, Oxford.

Banerjee, A., Dolado, J.J., Mestre, R. (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis* **19** 267-283.

Banerjee, A., Dolado, J.J., Hendry, D.F., Smith, G.W. (1986). Exploring equilibrium relationships in econometrics through static models: Some monte carlo evidence. *Oxford Bulletin of Economics and Statistics* **48** 253-277.

Bewley, R.A. (1979). The direct estimation of the equilibrium response in a linear model. *Economics Letters* **3** 357-361.

Boswijk, H.P. (1994). Testing for an unstable root in conditional and structural error correction models. *Journal of Econometrics* **63** 37-60.

Boswijk, H.P. (1995). Efficient inference on cointegration parameters in structural error correction models. *Journal of Econometrics* **69** 133-158.

Davidson, J.E.H., Hendry, D.F., Srba, F., Yeo, S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal* **88** 661-692.

Dickey, D.A., Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74** 427-431.

Dickey, D.A., Fuller, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49** 1057-1072.

Engle, R.F., Granger, C.W.J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica* **55** 251-276.

Ericsson, N.R. (2004). The ET interview: Professor David F. Hendry. *Econometric Theory* **20** 743-804.

Fuller, W.A. (1996). *Introduction to Statistical Time Series.* Wiley, 2nd edition, New York.

Hassler, U. (1999). (When) Should cointegrating regressions be detrended? The case of a German money demand function. *Empirical Economics* **24** 155-172.

Hassler, U. (2000). Cointegration testing in single error-correction equations in the presence of linear time trends. *Oxford Bulletin of Economics and Statistics* **62** 621-632.

Hendry, D.F (1979). Predictive failure and econometric modelling in macroeconomics: The transactions demand for money. In *Economic Modelling* (P. Ormerod, ed.), 217-242. Heinemann, London.

Hendry, D.F., Pagan, A., Sargan, J.D. (1984). Dynamic specification. In *Handbook of Econometrics, II* (Z. Griliches, M.D. Intriligator, eds.), 1023–1100. North-Holland, Amsterdam.

Hylleberg, S., Mizon, G.E. (1989). Cointegration and error correction mechanisms. *The Economic Journal* **99** 113-125.

Johansen, S. (1992). Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of Econometrics* **52** 389-402.

Lütkepohl, H. (2006). Structural vector autoregressive analysis for cointegrated variables. *Allgemeines Statistisches Archiv* **?** ?.

Pesaran, M.H., Shin, Y. (1998). An autoregressive distributed-lag modelling approach to cointegration analysis. In *Econometrics and Economic Theory in the 20th Century. The Ragnar Frisch Centennial Symposium* (S. Strøm, ed.), 371–413. Cambridge University Press, Cambridge.

Pesaran, M.H., Shin, Y., Smith, R.J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics* **16** 289-326.

Phillips, A.W. (1954). Stabilization policy in a closed economy. *Economic Journal* **64** 290–323.

Phillips, A.W. (1957). Stabilization policy and the time form of lagged responses. *Economic Journal* **67** 265–277.

Phillips, P.C.B. (1988). Reflections on econometric methodology. *The Economic Record* **64** 344–359.

Phillips, P.C.B., Durlauf, S.N. (1986). Multiple time series regressions with integrated processes. *Review of Economic Studies* **LIII** 473-495.

Phillips, P.C.B., Loretan, M. (1991). Estimating long-run economic equilibria. *Review of Economic Studies* **58** 407–436.

Phillips, P.C.B., Park, J.Y. (1988). Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association* **83** 111-115.

Sargan, J.D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology. In *Econometric Analysis for National Economic Planning* (R.E. Hart, G. Mills, J.K. Whittaker, eds.), 25–54. Butterworths, London.

Stock, J.H. (1987). Asymptotic properties of least-squares estimators of cointegrating vectors. *Econometrica* **55** 1035-1056.

WICKENS, M.R., BREUSCH, T.S (1988). Dynamic specification, the long-run and the estimation of transformed regression models. *The Economic Journal* **98** 189-205.

URBAIN, J.-P. (1992). On weak exogeneity in error correction models. *Oxford Bulletin of Economics and Statistics* **54** 187-207.

Uwe Hassler
Fachbereich Wirtschaftswissenschaften
Goethe-Universität Frankfurt
Gräfstr. 78, Uni-PF 76
60054 Frankfurt am Main
Deutschland
hassler@wiwi.uni-frankfurt.de

Jürgen Wolters
Fachbereich Wirtschaftswissenschaft
Freie Universität Berlin
Boltzmannstr. 20
14195 Berlin
Deutschland
wolters@wiwiss.fu-berlin.de