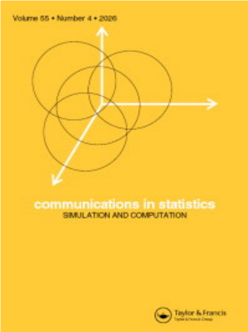

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: www.tandfonline.com/journals/lssp20

Estimating cointegrating models in the presence of additive outliers

Johan Lyhagen & Özge Baris-Tüzemen

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2025.2577404

Estimating cointegrating models in the presence of additive outliers

Johan Lyhagen and Özge Baris-Tüzemen

Department of Statistics, Uppsala University, Uppsala, Sweden

ABSTRACT

In this paper, we compare the finite-sample performance of the most widely used cointegration models in the presence of outliers: the Johansen estimator and the auto-regressive distributed lag (ARDL) framework, including both nonlinear and robust ARDL variants. Using Monte Carlo simulations across various outlier magnitudes, sample sizes, and both linear and nonlinear data-generating processes, we assess each model's robustness based on bias and mean squared error (MSE). The results indicate that the Johansen estimator is particularly sensitive to outliers, especially in small samples, while ARDL-based approaches, most notably the two-step nonlinear ARDL, demonstrate greater resilience and yield more accurate parameter estimates. These simulation results are further supported by an empirical application to U.S. industrial production and unemployment data.

ARTICLE HISTORY
Received 3 May 2023
Accepted 11 October 2025

KEYWORDS
Additive outliers; Cointegration; Error correction model; Monte Carlo simulations

1. Introduction

Most economic variables are nonstationary. Granger and Newbold (1974) argued that analyses involving nonstationary series may lead to spurious correlations, as illustrated by the famous "babies and storks" example. However, if a linear combination of two or more series with the same order of integration results in a series with a lower order of integration, the series are said to be cointegrated. Johansen (1995) cointegration estimator is one of the most widely used methods for estimating long-run relationships. Developed within the framework of the Vector Auto-Regression (VAR) model, the Johansen approach allows for the presence of multiple cointegrating vectors. This model also allows the testing of economic theories through cointegration and stochastic trends. In addition, with this approach, the adjustment parameters that explain how the system comes to equilibrium can be estimated and constraints on the cointegration relationships can be tested (Ankargren and Lyhagen 2019). Another widely used cointegration model in applied econometrics research is the Auto-Regressive Distributed Lag (ARDL) which has been introduced by Pesaran and Shin (1999), Pesaran et al. (2001). Considering the unrestricted error correction model, the ARDL approach can be applied regardless of whether the regressors are fully I(0) and/or fully I(1). As a single-equation model that is robust against the misspecifications, ARDL has an advantage over other equations.

Economic time series are often affected by external factors such as policy changes, strikes and measurement errors. These events appear as large residuals or outliers in econometric models (Nielsen, 2004, p. 249). According to Fox (1972) two types of outlier might occur in a time series.

CONTACT Özge Baris-Tüzemen ozge.baris-tuzemen@statistik.uu.se Department of Statistics, Uppsala University, PO Box 513, 752 20 Uppsala, Sweden.

The first of those is additive outliers (AOs) which are gross error of observation or recording error and affect single observation. The other one is innovation outliers (IOs), which correspond to single extreme shock and affect subsequent observations. Franses and Lucas (1998) point out that the reason why aberrant observations occur frequently in empirical applications is usually due to the approximate nature of the postulated model. Real life is more complicated than proposed statistical models. Therefore, atypical observations may not always be "bad" observations, on the contrary, they can disclose beneficial information about the functioning of the economic mechanism and constitute useful structures for model refinements. However, outliers should be accounted for in an appropriate way, otherwise, they might have a negative influence on regression and estimation of scale parameters, thus on the overall inference (Franses and Lucas 1998; Desgagné 2021). Franses and Haldrup (1994) find that the Johansen model may produce spurious cointegrating vectors in the presence of AOs since Johansen test is extremely sensitive to the deviation of error terms. Moreover, the Johansen model has a high tendency to produce greater variance and outliers (Maddala and Kim 1998, p. 173). To reduce the effects of outliers Lucas (1997) suggests the Student *t* pseudo-likelihood approach, which is a Johansen-type testing procedure as an alternative to cointegration analysis. Subsequently Franses and Lucas (1998) confirm that the new approach outperforms Johansen maximum likelihood test, if the residuals are leptokurtic. Although innovation outliers are often observed in economic time series, particularly at high frequencies (see Balke and Fomby 1994; Tolvi 2001), many studies emphasize that additive outliers can be more harmful in applied econometrics. For instance, Franses and Ghijsels (1999) show that ignoring additive outliers in financial return series can cause biased volatility forecasts. Similarly, Franses and Haldrup (1994) argue that additive outliers may pose a greater challenge than innovation outliers because the former can induce spurious stationarity. In particular, additive outliers may cause the Dickey–Fuller test to reject the unit root hypothesis too often and lead the Johansen test to detect too many cointegrating vectors. Nielsen (2004) further demonstrates that additive outliers increase the Type I error rate, while innovation outliers are comparatively easier to handle, as they can be modeled using unrestricted dummy variables. In contrast, additive outliers are more problematic in cointegrated VAR models due to their incompatibility with the reduced-rank structure. Nielsen also finds that pseudo-likelihood methods do not outperform Gaussian maximum likelihood when additive outliers are present, especially in higher-dimensional VAR models. Motivated by these findings, the present study focuses on the implications of additive outliers.

In this study, we consider the following cointegration models: the Johansen model, ARDL, quantile ARDL, nonlinear ARDL, and two-step nonlinear ARDL models. These approaches are widely used in empirical macroeconomic research to examine long-run equilibrium relationships among variables such as income, output, unemployment, stock prices, and interest rates. The Johansen model, which allows multiple cointegrating relationships and treats all variables as endogenous, is often favored for its forecasting benefits under assumptions of normality and linearity. However, such assumptions are frequently violated in real-world data, where structural breaks, nonlinearities, and outliers are common.[1] Nonlinear ARDL models have become particularly popular because they allow for asymmetric responses to positive and negative shocks, a feature aligned with the observed behavior of economic agents. Similarly, the ARDL model, which accounts for distributional and location asymmetries, have seen increasing use.

Although the effects of outliers on the Johansen maximum likelihood (ML) estimator are well-documented, ARDL-based cointegration models have not been evaluated extensively under such conditions. This paper addresses that gap by systematically comparing the performance of the Johansen model with various ARDL-based models in the presence of additive outliers. We present

[1]See, for example, Balke and Fomby (1994), Premaratne and Bera (2000), Doornik and Ooms (2005), Loperfido (2020), and Lee et al. (2022).

the first simulation-based comparison of the Johansen model and ARDL-based cointegration estimators including quantile ARDL, as well as single- and two-step NARDL models under data contamination from additive outliers. The results show that ARDL-based models tend to be more robust to outliers in many scenarios, providing valuable guidance for empirical applications. To do so, we conduct several Monte Carlo experiments, and compare bias and the mean squared error of the estimators. According to the simulation results, increasing the size of outlier leads to an increase in bias and mean squared error of parameter estimation. Apart from that, the efficiency of cointegration models depends on the type and magnitude of the outlier. The sample size also affects the performance of models. When considering the linear models, the ARDL model has the smallest mean squared error and bias while the Johansen model has the largest ones. Single- and two-step nonlinear ARDL models perform similar to each other, although there are exceptions according to the type of outlier. As the sample size becomes larger, the difference between those models becomes much smaller. However, two-step nonlinear ARDL model outperforms the single-step one. The main conclusion remains the same when errors are t-distributed.

The rest of the paper is organized as follows. Section 2 briefly describes the vector error correction model of the Johansen and ARDL-based models. Section 3 comprises a Monte Carlo study conduct to investigate the performance of the cointegration models. Section 4 contains an empirical analysis based on the U.S. macroeconomic data on industrial production and the unemployment rate. Finally, Sec. 5 concludes.

2. Cointegration models

In this section, we introduce the Johansen vector error correction model (VECM), the ARDL model, and ARDL kind of models such as nonlinear and quantile ARDL.

2.1. Johansen cointegration model

Let Y_t be a $p \times 1$ vector with elements $Y_t = y_t, x_{1,t}, ..., x_{p-1,t}$. The Johansen VECM can be written, following Johansen (1995):

$$\Delta Y_t = \alpha\beta^T Y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \varepsilon_t \tag{1}$$

where Δ denotes the first difference operator. The errors $\varepsilon_t \sim$ iid $\mathrm{N}(0, \Omega)$. The α and β are full column rank, $p \times r$ matrices, where $r < p$ and r is the cointegration rank. β are cointegration vectors and the matrix α is the adjustment coefficient that indicates the period needed to the long-run imbalances will be removed. Γ_i are $p \times p$ matrices that contain the short-run parameters, and $\Delta Y_t = Y_t - Y_{t-1}$. Johansen (1995) assumes that Y_t is an integrated order of one, hence, Eq. (1) is the I(1) model.

2.2. The ARDL model

The Johansen model requires all variables to be integrated of order one (I(1)). In contrast, the ARDL approach allows for the analysis of long-run relationships among variables regardless of the integration order of the regressors. The ARDL error correction model (ECM) facilitates the simultaneous investigation of both long-run relationships and short-run dynamics.

The ARDL modeling approach consists of three stages. In the first step, an unrestricted error correction model is estimated to test for the existence of a long-run association among the variables. In the second step, if cointegration is confirmed, the long-run parameters are estimated. Finally, the short-run dynamics and the error correction term are obtained. Unlike the Johansen

VECM, which is specified in vector form using Y_t to represent the system of variables, the ARDL model is expressed in scalar form, with the dependent variable y_t and its associated regressors $x_{i,t}$. The general ARDL (p, q) model in its unrestricted error correction form is given as follows (Pesaran et al. 2001):

$$\Delta y_t = \alpha\left(y_{t-1} + \sum_{i=1}^{p-1}\beta_i x_{i,t-1}\right) + \sum_{i=1}^{k-1}\delta_i \Delta y_{t-i} + \sum_{i=1}^{p-1}\sum_{j=0}^{k_i}\delta_{ij}\Delta x_{i,t-j} + \varepsilon_t \tag{2}$$

where β denotes the long-run parameters and δ represents the short-run coefficients. The parameter α is the error correction term, measuring the speed of adjustment back to the long-run equilibrium. The term ε_t is the error term.

Equation (2) shows that the ARDL model includes the lagged level of the dependent variable y_{t-1} and the lagged levels of the regressors $x_{i,t-1}$ in the long-run component. If a regressor is I(0), its level can be included directly in the long-run term without generating spurious regression problems. If a regressor is I(1), it can cointegrate with the dependent variable, preserving the validity of the long-run relationship. Therefore, the ARDL model can accommodate regressors with mixed integration orders while still testing for the existence of a stable long-run equilibrium among them (see Pesaran et al. 2001 for further discussion).

The error term in ARDL-based models is assumed to follow a white noise process with zero mean and constant variance. In contrast, the Johansen VAR model assumes that the error term is independently and identically distributed as a multivariate normal, iid $\mathcal{N}(0, \Omega)$, consistent with maximum likelihood estimation. Thus, while ARDL models assume white noise errors with constant variance, the Johansen model assumes i.i.d. multivariate normal errors.

2.3. Quantile ARDL model

The quantile ARDL combines the ARDL and quantile regression of Koenker and Bassett (1978). The model enables the estimation of short- and long-run connections between variables in different quantiles of the dependent variable's conditional distribution. It provides a more complete econometric framework as it considers heterogeneity between quantiles in both the short- and long-run relationships. In addition, when non-normality in the variables is ignored, it produces more robust evidence against the possibility of false acceptance of the non-cointegration hypothesis (Shahzad et al. 2021, p. 7). The quantile ARDL model is basically the same as the ARDL but the parameters depends on the quantile τ:

$$\begin{aligned}\Delta y_t = \alpha(\tau)&\left(y_{t-1} + \sum_{i=1}^{p-1}\beta_i(\tau)x_{i,t-1}\right)\\ &+\sum_{i=1}^{k-1}\delta_i(\tau)\Delta y_{t-i} + \sum_{i=1}^{p-1}\sum_{j=0}^{k_i}\delta_{ij}(\tau)\Delta x_{i,t-j} + \varepsilon_t(\tau)\end{aligned} \tag{3}$$

where $\tau (0 < \tau < 1)$ shows the quantile, while $\varepsilon_t(\tau)$ is the error term in period t.[2] Traditional regression model minimizes mean squared error. Contrary, the quantile model minimizes absolute deviation when τ is equal to 0.5.

2.4. Nonlinear ARDL model

Traditional cointegration models assume that positive and negative shocks have the same effect while examining the long-run relationship among the variables. However, it is believed that

[2]More detailed technical description of the QARDL approach, see the paper by Cho et al. (2015).

the reactions of economic units to positive and negative shocks are not the same, therefore, the asymmetric structure should be taken into account when investigating the relationship between the variables. In this context, Schorderet (2001) implemented the nonlinear long-run equilibrium based on the partial sum of the decomposition as part of the asymmetric cointegration among the variables. In the following year, Granger and Yoon (2002) suggested the concept of "Hidden Co-integration" where the cointegration relationship is described among the positive and negative components of the corresponding variables. More recently, Shin et al. (2014) introduced a nonlinear model within the framework of the ARDL approach. In the nonlinear ARDL (NARDL) regressor X_t is decomposed into its positive and negative partial sums, that is $X_t = X_0 + X_t^+ + X_t^-$, where $X_t^+ = \sum_{i=1}^t \Delta X_i^+ = \sum_{i=1}^t \max(\Delta X_i, 0)$, $X_t^- = \sum_{i=1}^t \Delta X_i^- = \sum_{i=1}^t \min(\Delta X_i, 0)$. Then the asymmetric long-run equilibrium relationship is expressed as: $y_t = \theta^+ X_t^+ + \theta^- X_t^- + e_t$. Where Y_t and X_t are scalar I (1) variables. The NARDL (p, q) based on ECM is as follows:

$$\Delta y_t = \alpha\left(y_{t-1} + \sum_{i=1}^{p-1}\beta_i^+ x_{i,t-1}^+ + \sum_{i=1}^{p-1}\beta_i^- x_{i,t-1}^-\right) + \sum_{i=1}^{k-1}\delta_i \Delta y_{t-i} + \sum_{i=1}^{p-1}\sum_{j=0}^{k_i}\left(\delta_{ij}^+ \Delta x_{i,t-j}^+ + \delta_{ij}^- \Delta x_{i,t-j}^-\right) + \varepsilon_t \tag{4}$$

where α is the nonlinear error correction term.

2.5. Two-step nonlinear ARDL model

The two-step nonlinear ARDL is basically the same as NARDL but the short- and long-run parameters are estimated in two step unlike the single-equation ARDL based models. In the NARDL model, it is stated that the positive and negative cumulative sums of the independent variables are dominated by the asymptotically co-linear deterministic time trend. According to Cho et al. (2019), this co-linear trend generate an asymptotic singularity problem. That is an important obstacle to the development of the asymptotic theory for the single-step estimation framework. Therefore, the NARDL model does not follow the asymptotic theory, the model carries out Monte Carlo simulations to validate the properties of the single-step OLS estimator in finite samples instead (Cho et al. 2019). To overcome the singularity problem in the single-step NARDL model, Cho et al. (2019) propose a two-step NARDL model. In the first stage of this model, Phillips and Hansen 1990) fully modified OLS (FM-OLS) model is estimated to obtain the long-run parameters. In the second stage, the OLS model is estimated for the short-run parameters. Contrary to the OLS model, the FM-OLS model provides robust results against serial correlation and potential endogeneity between the regressors (Wagner and Hong 2016). Also, considering the consistency of the long-run coefficients in the first step, the error correction term can be processed as known in the second step because the OLS provides a consistent and asymptotically normal estimator for the short-run parameters (Cho et al. 2019, p. 3).

3. Monte Carlo simulations design and results

In this section, first, we conduct a Monte Carlo experiment to compare the performance of models which are described in Sec. 2. Then we report Monte Carlo simulation results.

3.1. Data generating process (DGP)

As for the data-generating process (DGP), we adopt the two-step nonlinear ARDL specification proposed by Cho et al. (2015) as our benchmark. This DGP enables a direct comparison between linear and nonlinear models in terms of parameter estimates. Moreover, Cho et al. (2015) address the singularity issues inherent in single-step nonlinear ARDL estimation and employ a fully modified OLS model to consistently estimate long-run relationships. Given these advantages, their framework is well suited to the main objective of our study, which is to evaluate model performance under additive outlier contamination. Therefore, we first employ the DGP from Cho et al. (2015), and subsequently modify it in various ways to suit our simulation design. The baseline NARDL $(1, 0)$ DGP is specified as follows:

$$\Delta y_t = \gamma + \rho u_{t-1} + \phi \Delta y_{t-1} + \pi^+ \Delta x_t^+ + \pi^- \Delta x_t^- + e_t \tag{5}$$

where $u_{t-1} := y_{t-1} - \alpha - \beta^+ x_{t-1}^+ - \beta^- x_{t-1}^-$, $\Delta x_t := \kappa \Delta x_{t-1} + \sqrt{1-\kappa^2} v_t$, and $(e_t, v_t)' \sim$ iid $\mathrm{N}(0_2, I_2)$. The parameters we use; $(\alpha, \beta^+, \beta^-, \gamma, \rho, \phi, \pi^+, \pi^-, \kappa) = (0, 2, 1, 0, -2/3, \phi, 1, 1/2, 1/2)$. We repeat the DGP 10000 times considering the small sample sizes 50, 100, 200, 400, which are typical for macroeconomic variables. We generate data with $T + 50$ observations for each series, and then discard the first 50 observations to be able to eliminate initial effects. Cho et al. (2015) suggest varying the autoregressive parameter ϕ to induce different degrees of serial correlation. Following this approach, we consider three values for ϕ: -0.5, 0.0, and 0.5, which reflect one of the main assumptions of the ARDL model (see Pesaran et al. 2001 for a discussion on this point). To evaluate the performance of the estimators, we compute the sample bias and the mean squared error (MSE) of the parameter estimates. Two data-generating processes (DGPs) are specified for comparison: one linear and one nonlinear. In the linear DGP, we use the mean values of the parameters β and π from Eq. (5) for the relevant components. Finally, for the QARDL model, we use $\tau = 0.5$, which corresponds to the median and serves as a more robust estimator of central tendency. The choice of $\tau = 0.5$ in this study aligns with its well-documented robustness properties in the quantile regression literature. Specifically, it minimizes the sum of absolute deviations rather than squared deviations, making it less sensitive to outliers compared to mean regression. This robustness makes the median quantile estimator more reliable in the presence of heavy-tailed or non-normal error distributions, which are common in macroeconomic data. For a technical discussion of the asymptotic behavior and efficiency of the median quantile estimator, see Koenker and Bassett (1978).

To assess the robustness of the different cointegration models under both linear and nonlinear settings, we introduce various types of contamination into the data.[3] Specifically, we contaminate the original series y_t and x_t with 5% additive outliers, defined as follows:[4]

I. **Additive outliers in x_t only:** $x_t^* = x_t + \theta u_t$, with $y_t = f(x_t^*)$.
II. **Measurement error in x_t:** $y_t = f(x_t)$, but estimation is performed using contaminated x_t^*.
III. **Additive outliers in y_t:** $y_t^* = y_t + \theta u_t$, and $y_t = f(x_t, y_{t-1}^*)$.
IV. **Measurement error in y_t:** $y_t = f(x_t, y_{t-1})$, but observed values are y_t^*.

In all cases, u_t follows an i.i.d. process drawn from a uniform distribution on the unit circle, and θ is set to four values: $\theta \in 0, 3, 5, 7$. This setup allows us to examine model robustness across varying outlier magnitudes.

Additionally, we examine the effect of non-normal error distributions. After generating a standard Gaussian process, we replace the error terms v_t and e_t with draws from a t-distribution with 5 degrees of freedom. We consider three scenarios:

- **Case 1:** $e_t \sim t_5$, $v_t \sim$ iid $\mathcal{N}(0, 1)$

[3]For the choice of contamination rate and magnitudes, we are guided by Lucas (1997), Franses and Lucas (1998), and Franses et al. (1998), which are among the primary studies examining the effects of additive outliers on cointegration models.
[4]The paper differs Franses and Lucas (1998) for the case (ii) and (iii).

- **Case 2:** $v_t \sim t_5$, $e_t \sim \text{iid } \mathcal{N}(0,1)$
- **Case 3:** $e_t \sim t_5$, $v_t \sim t_5$

To ensure comparability with the normally distributed errors, we use standardized t-distributions, and the outlier magnitude is further benchmarked at five times the residual standard deviation.

To facilitate comparison across models, we rewrite the parameters in the Johansen model in conditional form (see Eq. (8.3) in Johansen (1995), p. 122), allowing structural parallels with the ARDL-based specifications. For the two-step nonlinear ARDL model, long-run parameters are estimated following Cho et al. (2019), p. 14. In all models, we extract the product of the long-run coefficient vector and the adjustment coefficient (i.e. $\beta \times \rho$) to ensure comparability, especially since normalization in the Johansen framework can yield disproportionately large coefficients. For the ARDL, quantile ARDL, and single-step nonlinear ARDL models, this transformation is standard and supported in R packages. For the Johansen and two-step nonlinear ARDL models, we compute β parameters manually within the current study.

3.2. Results

In this subsection we report simulations results. The estimators are: the Johansen model (ML), auto-regressive distributed lag (ARDL), quantile ARDL (QARDL), single-step nonlinear ARDL (NARDL), and two-step NARDL (2NARDL) model. We are unable to present the all simulations' results herein due to space restrictions, but we summarize the most interesting findings.[5] We also do not report the results of the constants, α and γ.

In Tables 1 and 2, we report the linear DGP's MSE results for case (i), when θ is equal to 0 and 3, respectively. As seen in Table 1, the ML estimator has the largest MSE, whereas the ARDL model consistently has the smallest across all sample sizes. Among the nonlinear estimators, the 2NARDL performs better than the NARDL, although the difference is slight. As expected, the MSE of all models decreases as the sample size increases. The bias results follow a similar pattern to the MSE results for all models. From Table 2, it can be seen that when the data contain outliers in the x-direction, the MSEs of all models, except NARDL, increase. In particular, the MSEs of the β parameters become much larger as the sample size increases. The ARDL model performs best among the models considered. According to the results in Table 2, the MSE values of the NARDL model decrease as the magnitude of the outlier increases, which may be due to the small sample size. When T is as large as 5000, NARDL's MSE converges to that of the ARDL. Summarizing the results for case (i): both linear and nonlinear DGP simulations follow a similar pattern when considering the magnitude of outliers. However, it is important to emphasize that ML, ARDL, and QARDL exhibit much larger MSEs in the nonlinear DGP case. This holds for all cases and distributions and the MSEs increase with larger T.

Table 3 reports the results of the nonlinear DGP for case (ii) with a standard deviation of residuals equal to 5. As expected, linear models exhibit much larger MSE and bias in the nonlinear DGP setting. In this case, the ARDL models generally have smaller MSE and bias compared to the other linear models. Both the NARDL and 2NARDL models are highly affected by outliers, and their MSEs increase as T increases. However, except for a few parameters, the 2NARDL yields smaller MSE values. Table 4 presents the results of the linear DGP for the same case and outlier magnitude as in Table 3. While the MSEs of the ARDL and QARDL models increase as θ increases for all sample sizes, the performance of the ML model depends on both the sample size and the magnitude of θ. When θ and the sample size are small, no significant change is observed; however, as θ becomes large (e.g. $\theta = 9$), the small-sample problem disappears, and the ML

[5]All results are available on request.

Table 1. Monte Carlo simulation results: finite sample mean squared error (MSE) of the estimators (MSE $\times$ 100).

		50					100					200					400				
ϕ	|	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL
−0.5	ρ	2.489	1.011	1.540	1.491	1.491	0.827	0.395	0.613	0.476	0.514	0.349	0.176	0.274	0.191	0.201	0.159	0.082	0.130	0.086	0.089
	$\rho\times\beta^+$	–	–	–	2.443	2.269	–	–	–	0.900	0.898	–	–	–	0.398	0.399	–	–	–	0.186	0.186
	$\rho\times\beta^-$	–	–	–	2.421	2.254	–	–	–	0.906	0.898	–	–	–	0.397	0.398	–	–	–	0.186	0.187
	$\rho\times\beta$	4.742	1.881	2.876	–	–	1.725	0.814	1.256	–	–	0.759	0.379	0.591	–	–	0.351	0.181	0.286	–	–
	ϕ	1.171	0.606	0.922	0.732	0.934	0.422	0.252	0.390	0.273	0.329	0.186	0.117	0.181	0.120	0.132	0.086	0.055	0.086	0.055	0.058
	π^+	–	–	–	12.203	12.809	–	–	–	4.726	4.934	–	–	–	2.216	2.253	–	–	–	1.030	1.039
	π^-	–	–	–	12.001	12.740	–	–	–	4.726	4.989	–	–	–	2.140	2.169	–	–	–	1.059	1.064
	π	3.396	2.877	4.418	–	–	1.499	1.291	2.065	–	–	0.709	0.619	0.983	–	–	0.351	0.308	0.476	–	–
0	ρ	3.769	1.542	2.259	1.981	1.749	1.368	0.665	1.010	0.762	0.714	0.582	0.306	0.473	0.329	0.319	0.272	0.148	0.231	0.154	0.152
	$\rho\times\beta^+$	–	–	–	4.182	3.632	–	–	–	1.648	1.554	–	–	–	0.730	0.707	–	–	–	0.343	0.337
	$\rho\times\beta^-$	–	–	–	4.180	3.648	–	–	–	1.652	1.557	–	–	–	0.728	0.706	–	–	–	0.343	0.337
	$\rho\times\beta$	8.219	3.296	4.870	–	–	3.033	1.467	2.225	–	–	1.305	0.683	1.052	–	–	0.609	0.331	0.517	–	–
	ϕ	1.657	0.848	1.374	0.941	0.859	0.730	0.403	0.629	0.424	0.402	0.344	0.197	0.311	0.201	0.198	0.167	0.094	0.150	0.095	0.094
	π^+	–	–	–	12.378	11.816	–	–	–	4.791	4.688	–	–	–	2.248	2.234	–	–	–	1.039	1.035
	π^-	–	–	–	12.181	11.866	–	–	–	4.796	4.727	–	–	–	2.160	2.137	–	–	–	1.066	1.060
	π	3.417	2.990	4.529	–	–	1.500	1.321	2.091	–	–	0.709	0.635	1.008	–	–	0.351	0.314	0.486	–	–
0.5	ρ	1.525	0.975	1.489	1.149	1.034	0.615	0.443	0.678	0.482	0.457	0.280	0.207	0.322	0.216	0.211	0.136	0.102	0.160	0.104	0.103
	$\rho\times\beta^+$	–	–	–	3.316	2.952	–	–	–	1.197	1.183	–	–	–	0.514	0.508	–	–	–	0.240	0.240
	$\rho\times\beta^-$	–	–	–	3.338	2.977	–	–	–	1.193	1.185	–	–	–	0.513	0.508	–	–	–	0.240	0.240
	$\rho\times\beta$	3.955	2.406	3.670	–	–	1.465	1.043	1.594	–	–	0.654	0.477	0.743	–	–	0.312	0.231	0.363	–	–
	ϕ	1.048	0.636	1.008	0.692	0.692	0.502	0.301	0.468	0.313	0.313	0.241	0.143	0.224	0.146	0.147	0.120	0.069	0.109	0.069	0.069
	π^+	–	–	–	12.234	11.833	–	–	–	4.716	4.630	–	–	–	2.216	2.206	–	–	–	1.023	1.020
	π^-	–	–	–	12.143	11.898	–	–	–	4.749	4.670	–	–	–	2.137	2.117	–	–	–	1.048	1.041
	π	3.416	2.919	4.354	–	–	1.506	1.273	2.014	–	–	0.710	0.615	0.970	–	–	0.350	0.299	0.463	–	–

Note. Evaluation of the models' performance when DGP is linear, the case (i) and $\theta = 0$. 50, 100, 200, and 400 indicate the sample sizes. ML represents the Johansen model, while ARDL is the auto-regressive distributed lag model. QARDL is the quantile ARDL, NARDL is the single-step nonlinear ARDL, and 2NARDL is the two-step NARDL.

Table 2. Monte Carlo simulation results: finite sample mean squared error (MSE) of the estimators (MSE × 100).

		50					100					200					400				
ϕ		ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL
−0.5	ρ	12.572	9.582	9.951	1.357	3.518	22.984	18.660	18.782	0.439	6.189	34.957	31.019	31.020	0.177	11.257	41.465	39.649	39.664	0.080	18.014
	$\rho\times\beta^{+}$	–	–	–	2.237	7.885	–	–	–	0.833	12.576	–	–	–	0.370	22.740	–	–	–	0.173	37.624
	$\rho\times\beta^{-}$	–	–	–	2.258	12.800	–	–	–	0.838	19.084	–	–	–	0.368	29.839	–	–	–	0.173	44.320
	$\rho\times\beta$	23.028	16.406	17.233	–	–	45.792	36.089	36.358	–	–	73.146	63.893	63.929	–	–	89.359	84.660	84.709	–	–
	ϕ	1.938	1.348	1.748	0.683	1.112	2.536	1.180	1.414	0.258	0.593	3.727	1.193	1.330	0.115	0.474	4.332	1.130	1.229	0.053	0.487
	π^{+}	–	–	–	8.564	11.964	–	–	–	3.431	5.987	–	–	–	1.582	4.237	–	–	–	0.750	4.292
	π^{-}	–	–	–	12.103	17.501	–	–	–	4.740	9.654	–	–	–	2.145	8.342	–	–	–	1.055	9.872
	π	4.197	5.346	7.122	–	–	2.016	6.413	7.677	–	–	1.035	9.799	10.650	–	–	0.495	12.549	13.091	–	–
0	ρ	15.397	11.430	11.884	1.833	5.382	27.050	22.460	22.565	0.707	10.228	37.880	34.506	34.509	0.308	17.058	42.673	41.307	41.318	0.145	24.500
	$\rho\times\beta^{+}$	–	–	–	3.851	11.108	–	–	–	1.529	20.607	–	–	–	0.684	35.433	–	–	–	0.324	52.492
	$\rho\times\beta^{-}$	–	–	–	3.951	16.455	–	–	–	1.547	27.516	–	–	–	0.686	42.199	–	–	–	0.324	58.105
	$\rho\times\beta$	33.270	23.456	24.498	–	–	58.837	47.907	48.173	–	–	83.274	75.130	75.148	–	–	94.589	91.176	91.215	–	–
	ϕ	1.906	1.326	1.959	0.893	1.224	1.616	0.786	1.126	0.406	0.715	1.822	0.513	0.677	0.193	0.449	1.814	0.387	0.470	0.091	0.332
	π^{+}	–	–	–	8.689	10.836	–	–	–	3.464	5.178	–	–	–	1.596	3.239	–	–	–	0.754	2.678
	π^{-}	–	–	–	12.299	15.725	–	–	–	4.798	7.940	–	–	–	2.164	5.961	–	–	–	1.062	5.799
	π	4.105	4.295	6.108	–	–	2.022	3.855	4.831	–	–	1.153	4.770	5.369	–	–	0.715	5.366	5.675	–	–
0.5	ρ	9.757	8.402	8.737	1.068	3.796	20.562	18.448	18.540	0.445	7.389	33.426	31.320	31.330	0.201	12.985	41.041	39.959	39.970	0.098	19.992
	$\rho\times\beta^{+}$	–	–	–	2.976	7.475	–	–	–	1.088	14.058	–	–	–	0.476	26.080	–	–	–	0.225	42.014
	$\rho\times\beta^{-}$	–	–	–	3.234	12.033	–	–	–	1.125	20.713	–	–	–	0.485	33.231	–	–	–	0.227	48.497
	$\rho\times\beta$	25.304	20.112	20.969	–	–	48.941	42.415	42.724	–	–	77.076	70.901	70.949	–	–	93.548	90.096	90.135	–	–
	ϕ	1.705	1.007	1.481	0.647	0.941	2.622	0.844	1.108	0.291	0.592	4.241	0.994	1.158	0.135	0.523	5.301	1.155	1.257	0.064	0.613
	π^{+}	–	–	–	8.603	11.210	–	–	–	3.419	5.419	–	–	–	1.585	3.133	–	–	–	0.746	2.264
	π^{-}	–	–	–	12.195	16.410	–	–	–	4.753	8.181	–	–	–	2.143	5.471	–	–	–	1.044	4.748
	π	4.683	4.247	6.072	–	–	2.687	3.287	4.422	–	–	1.766	3.851	4.448	–	–	1.275	4.333	4.715	–	–

Note. Evaluation of the models' performance when DGP is linear, the case (i) and $\theta = 3$. 50, 100, 200, and 400 indicate the sample sizes. ML represents the Johansen model, while ARDL is the auto-regressive distributed lag model. QARDL is the quantile ARDL, NARDL is the single-step nonlinear ARDL, and 2NARDL is the two-step NARDL.

Table 3. Monte Carlo simulation results: finite sample mean squared error (MSE) of the estimators (MSE × 100).

		50					100					200					400				
ϕ		ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL
−0.5	ρ	34.269	31.350	31.527	7.646	4.502	41.148	39.532	39.684	13.827	6.482	43.492	42.869	42.933	22.236	10.493	44.199	44.011	44.031	30.245	16.348
	$\rho \times \beta^+$	–	–	–	40.031	24.837	–	–	–	64.573	32.950	–	–	–	96.626	48.165	–	–	–	126.236	70.656
	$\rho \times \beta^-$	–	–	–	5.987	5.172	–	–	–	9.383	6.015	–	–	–	17.862	9.336	–	–	–	26.816	14.868
	$\rho \times \beta$	59.067	54.878	54.697	–	–	77.227	74.132	74.026	–	–	88.932	86.830	86.644	–	–	94.849	93.744	93.616	–	–
	ϕ	4.409	2.836	3.408	1.460	1.369	4.165	2.064	2.512	1.005	0.681	3.868	1.448	1.803	0.813	0.440	3.548	1.008	1.302	0.713	0.368
	π^+	–	–	–	24.836	29.163	–	–	–	18.927	22.173	–	–	–	15.534	18.464	–	–	–	13.635	16.054
	π^-	–	–	–	28.708	25.018	–	–	–	23.982	17.638	–	–	–	26.850	18.276	–	–	–	31.268	21.641
	π	6.835	9.399	13.134	–	–	3.595	5.236	8.671	–	–	2.240	3.028	6.407	–	–	1.712	2.025	5.276	–	–
0	ρ	35.097	32.277	32.493	8.613	6.809	41.952	40.522	40.608	17.187	10.618	43.900	43.377	43.409	26.718	16.165	44.327	44.183	44.190	34.140	22.870
	$\rho \times \beta^+$	–	–	–	49.457	33.236	–	–	–	81.960	48.504	–	–	–	116.396	70.001	–	–	–	142.392	95.759
	$\rho \times \beta^-$	–	–	–	8.425	6.640	–	–	–	15.004	9.478	–	–	–	24.708	14.754	–	–	–	32.658	21.481
	$\rho \times \beta$	72.670	68.314	68.351	–	–	87.420	86.581	86.181	–	–	94.903	94.739	94.438	–	–	97.963	97.961	97.735	–	–
	ϕ	3.304	1.914	2.626	1.761	1.621	2.010	1.195	1.432	1.125	1.019	1.371	0.961	0.917	0.858	0.756	0.979	0.917	0.705	0.780	0.674
	π^+	–	–	–	26.315	28.773	–	–	–	21.389	23.172	–	–	–	19.156	20.623	–	–	–	18.232	19.251
	π^-	–	–	–	24.287	23.140	–	–	–	17.129	15.257	–	–	–	16.954	14.093	–	–	–	17.904	14.817
	π	7.877	7.581	10.111	–	–	5.083	3.463	4.839	–	–	3.954	1.603	2.509	–	–	3.507	0.850	1.457	–	–
0.5	ρ	30.621	28.446	28.450	6.005	5.121	40.187	38.878	38.904	13.250	7.972	43.518	42.948	42.965	22.574	12.760	44.289	44.084	44.088	30.923	19.141
	$\rho \times \beta^+$	–	–	–	44.638	24.822	–	–	–	72.488	36.988	–	–	–	105.634	56.307	–	–	–	133.875	81.276
	$\rho \times \beta^-$	–	–	–	8.837	4.731	–	–	–	13.815	6.693	–	–	–	22.621	11.259	–	–	–	30.722	17.628
	$\rho \times \beta$	79.324	70.795	70.887	–	–	96.148	90.879	90.630	–	–	100.687	98.226	97.951	–	–	100.693	100.060	99.885	–	–
	ϕ	4.232	1.415	2.126	0.995	1.038	4.370	0.933	1.449	0.519	0.516	4.493	0.804	1.169	0.436	0.349	4.395	0.707	1.010	0.486	0.330
	π^+	–	–	–	26.047	27.408	–	–	–	21.661	22.841	–	–	–	19.505	20.769	–	–	–	18.685	19.643
	π^-	–	–	–	23.239	24.133	–	–	–	15.122	14.917	–	–	–	14.069	12.593	–	–	–	14.768	12.534
	π	9.354	8.472	11.315	–	–	6.174	3.950	5.414	–	–	4.865	1.889	2.537	–	–	4.301	1.050	1.260	–	–

Note. Evaluation of the models' performance when DGP is nonlinear, the case (ii) and $\theta = 5$. 50, 100, 200, and 400 indicate the sample sizes. ML represents the Johansen model, while ARDL is the auto-regressive distributed lag model. QARDL is the quantile ARDL, NARDL is the single-step nonlinear ARDL, and 2NARDL is the two-step NARDL.

Table 4. Monte Carlo simulation results: finite sample mean squared error (MSE) of the estimators (MSE $\times$ 100).

		50					100					200					400				
ϕ		ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL
−0.5	ρ	2.073	2.438	2.176	5.108	2.523	0.739	1.747	1.008	9.611	1.748	0.326	1.404	0.558	17.192	1.498	0.157	1.266	0.374	25.736	1.411
	$\rho\times\beta^+$	–	–	–	17.627	7.006	–	–	–	27.995	4.781	–	–	–	44.457	3.734	–	–	–	62.206	3.340
	$\rho\times\beta^-$	–	–	–	7.946	6.953	–	–	–	15.768	4.755	–	–	–	32.771	3.716	–	–	–	52.941	3.329
	$\rho\times\beta$	3.967	5.918	4.708	–	–	1.545	4.159	2.290	–	–	0.714	3.269	1.277	–	–	0.352	2.899	0.853	–	–
	ϕ	1.303	0.955	1.146	1.287	1.156	0.542	0.435	0.509	0.823	0.471	0.258	0.209	0.241	0.706	0.210	0.134	0.102	0.118	0.701	0.100
	π^+	–	–	–	16.968	23.579	–	–	–	11.599	18.139	–	–	–	8.865	15.972	–	–	–	7.351	15.020
	π^-	–	–	–	22.862	19.262	–	–	–	18.022	10.105	–	–	–	21.023	7.279	–	–	–	26.431	6.111
	π	3.967	6.357	7.589	–	–	2.440	4.296	4.437	–	–	1.811	3.302	2.934	–	–	1.526	2.823	2.127	–	–
0	ρ	2.708	3.745	3.318	5.987	4.243	1.062	3.438	2.063	12.872	3.790	0.556	3.288	1.468	22.171	3.627	0.362	3.284	1.207	30.623	3.613
	$\rho\times\beta^+$	–	–	–	22.575	10.358	–	–	–	37.631	8.915	–	–	–	56.730	8.330	–	–	–	73.409	8.199
	$\rho\times\beta^-$	–	–	–	11.957	10.316	–	–	–	25.608	8.887	–	–	–	46.727	8.310	–	–	–	66.440	8.187
	$\rho\times\beta$	5.841	9.618	7.955	–	–	2.359	8.337	4.896	–	–	1.253	7.687	3.410	–	–	0.820	7.528	2.765	–	–
	ϕ	1.841	1.453	1.716	1.588	1.329	0.933	0.840	0.885	0.925	0.726	0.523	0.533	0.478	0.633	0.427	0.344	0.401	0.293	0.526	0.295
	π^+	–	–	–	17.905	22.361	–	–	–	13.043	17.714	–	–	–	10.937	15.943	–	–	–	10.045	15.163
	π^-	–	–	–	20.690	18.647	–	–	–	13.849	10.051	–	–	–	13.951	7.102	–	–	–	15.474	5.831
	π	4.198	6.480	7.693	–	–	2.708	4.471	4.424	–	–	2.080	3.482	2.872	–	–	1.807	3.022	2.080	–	–
0.5	ρ	1.436	2.598	2.186	4.090	3.107	0.643	2.305	1.321	9.586	2.624	0.355	2.183	0.906	18.051	2.424	0.229	2.143	0.717	26.920	2.348
	$\rho\times\beta^+$	–	–	–	21.241	7.053	–	–	–	33.482	5.756	–	–	–	50.826	5.364	–	–	–	67.914	5.226
	$\rho\times\beta^-$	–	–	–	11.016	7.028	–	–	–	21.344	5.743	–	–	–	39.978	5.354	–	–	–	59.782	5.221
	$\rho\times\beta$	3.663	7.564	5.959	–	–	1.551	6.002	3.351	–	–	0.832	5.298	2.183	–	–	0.529	5.010	1.675	–	–
	ϕ	1.140	0.871	1.161	0.951	0.921	0.590	0.409	0.541	0.502	0.418	0.317	0.200	0.264	0.446	0.203	0.200	0.105	0.136	0.547	0.103
	π^+	–	–	–	17.996	21.436	–	–	–	13.592	17.582	–	–	–	11.513	16.232	–	–	–	10.652	15.662
	π^-	–	–	–	20.181	19.186	–	–	–	12.490	9.869	–	–	–	11.570	6.431	–	–	–	12.662	4.971
	π	4.137	6.611	7.638	–	–	2.660	4.791	4.629	–	–	2.051	3.869	3.125	–	–	1.788	3.445	2.339	–	–

Note. Evaluation of the models' performance when DGP is linear, the case (ii) and $\theta = 5$. 50, 100, 200, and 400 indicate the sample sizes. ML represents the Johansen model, while ARDL is the auto-regressive distributed lag model. QARDL is the quantile ARDL, NARDL is the single-step nonlinear ARDL, and 2NARDL is the two-step NARDL.

model performs better. Similar to the nonlinear DGP, NARDL and 2NARDL have very large MSEs. However, unlike in Table 3, the MSE of 2NARDL decreases as the sample size increases, whereas this is not the case for NARDL. For case (ii), the large MSEs are mostly related to high bias across all models.

Finally, Tables 5 and 6 present the MSE results of the linear and nonlinear DGP for case (iv), where $\theta = 5$, respectively. When the data include additive outliers in the y-direction, as seen from Table 5, the ML model has the largest MSE for $T = 50$ when considering the same parameters, although this is not true for all sample sizes. The performance of the ARDL and QARDL models depends on the parameter in question: for some parameters, the ARDL has the smaller MSE, while for others the QARDL achieves the smallest value. However, for larger sample sizes, the QARDL generally performs better than the ARDL. The effect of outliers on the nonlinear ARDL models is almost the same: both are more vulnerable in small samples, and there is only a slight difference between their performance. Nevertheless, one may say that the 2NARDL performs better than the NARDL against outliers, especially for small sample sizes. These results are the same for these models in the case of the nonlinear DGP, as seen in Table 6.

Considering all the simulation results, one point should be noted: nonlinear models yield almost the same results for both linear and nonlinear DGPs. However, this does not hold for linear models. Given the nonlinear DGP, linear models appear to deviate significantly from the true model.

After generating data with Gaussian errors, we conduct Monte Carlo simulations using t-distribution and compare the results with the evidence under normal distribution. For both linear and nonlinear DGP, for all cases, we find that the results of both distributions are very similar to each other, and the main conclusion still holds. Therefore, the results are not reported herein.

4. Empirical example

In this section we provide an empirical example in order to illustrate the behavior of cointegration models. To do so, we investigate the Okun's Law, which assumes that there is a negative relationship between unemployment rate and economic growth. Due to its socioeconomic effects, unemployment is an important problem worldwide. Many developing economies, in particular, have been suffering from high unemployment rates for years. Beyond its social and economic consequences, identifying the fundamental causes of unemployment remains a crucial challenge. However, it is widely accepted that higher output growth increases employment and reduces unemployment Kreishan (2011). In the applied economics literature, the majority of studies have usually focused on linear paradigm between relevant variables (see, for some examples, Okun 1963; Sögner and Stiassny 2002; Dunsch 2017). For last two decades researcher have been pointing out that cyclical upturns and downturns do not have symmetrical influences on unemployment. It is believed that if employers lay off a certain number of employee during recessions, they will not hire exactly the same amount after a positive shock of same magnitude when considering high hiring process costs and hesitation of discouraged workers (Lee 2000; Schorderet 2003; Lang and de Peretti 2009; Shin et al. 2014).

Balke and Fomby (1994) analyze 15 post-World War II U.S. macroeconomic time series for the presence of additive and innovation outliers using linear autoregressive models. Their dataset spans 1947Q1 to 1992Q4 (with minor exceptions for some variables), and they find that many U.S. macroeconomic series exhibit outliers, especially during recessions or the early stages of recovery. Dijk et al. (1999) use a slightly adjusted version of the same dataset and demonstrate that neglecting additive outliers can negatively affect both the asymptotic size and power of

Table 5. Monte Carlo simulation results: finite sample mean squared error (MSE) of the estimators (MSE $\times$ 100).

		50					100					200					400				
ϕ	|	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL
−0.5	ρ	3.611	2.761	2.265	3.272	2.460	1.271	1.740	1.008	1.826	1.459	0.564	1.308	0.526	1.320	1.147	0.275	1.133	0.342	1.135	1.049
	$\rho\times\beta^+$	–	–	–	5.871	6.122	–	–	–	3.599	3.845	–	–	–	2.766	2.815	–	–	–	2.470	2.479
	$\rho\times\beta^-$	–	–	–	5.863	6.134	–	–	–	3.595	3.875	–	–	–	2.768	2.838	–	–	–	2.471	2.492
	$\rho\times\beta$	6.833	5.395	4.334	–	–	2.617	3.663	2.119	–	–	1.211	2.836	1.135	–	–	0.604	2.509	0.759	–	–
	ϕ	2.770	1.755	1.512	1.836	2.371	1.384	1.062	0.695	1.035	1.326	0.830	0.696	0.361	0.676	0.796	0.590	0.527	0.201	0.516	0.569
	π^+	–	–	–	21.234	21.267	–	–	–	8.921	8.804	–	–	–	4.088	4.034	–	–	–	2.015	1.990
	π^-	–	–	–	21.920	22.111	–	–	–	8.765	8.641	–	–	–	4.194	4.146	–	–	–	2.034	2.004
	π	5.255	5.444	5.750	–	–	2.218	2.475	2.520	–	–	1.048	1.245	1.204	–	–	0.509	0.654	0.575	–	–
0	ρ	5.931	2.150	2.632	2.269	2.162	2.729	1.442	1.485	1.363	1.390	1.504	1.202	1.025	1.131	1.168	0.935	1.135	0.822	1.093	1.115
	$\rho\times\beta^+$	–	–	–	5.058	4.999	–	–	–	3.051	3.234	–	–	–	2.536	2.671	–	–	–	2.460	2.531
	$\rho\times\beta^-$	–	–	–	5.085	5.019	–	–	–	3.048	3.261	–	–	–	2.538	2.692	–	–	–	2.461	2.544
	$\rho\times\beta$	13.050	4.703	5.731	–	–	6.078	3.232	3.303	–	–	3.367	2.698	2.296	–	–	2.100	2.556	1.849	–	–
	ϕ	1.688	1.629	1.897	1.752	1.712	0.748	1.016	1.040	1.030	1.060	0.393	0.737	0.650	0.731	0.760	0.224	0.603	0.465	0.599	0.613
	π^+	–	–	–	17.848	16.989	–	–	–	7.265	7.063	–	–	–	3.254	3.208	–	–	–	1.550	1.541
	π^-	–	–	–	18.277	17.722	–	–	–	7.135	6.906	–	–	–	3.300	3.268	–	–	–	1.554	1.542
	π	4.989	4.500	5.454	–	–	2.128	1.957	2.380	–	–	1.016	0.947	1.165	–	–	0.481	0.466	0.552	–	–
0.5	ρ	2.599	1.178	1.459	1.333	1.248	1.171	0.544	0.675	0.569	0.555	0.612	0.263	0.327	0.263	0.266	0.361	0.139	0.174	0.137	0.141
	$\rho\times\beta^+$	–	–	–	3.949	3.484	–	–	–	1.426	1.375	–	–	–	0.620	0.612	–	–	–	0.315	0.314
	$\rho\times\beta^-$	–	–	–	4.004	3.436	–	–	–	1.425	1.363	–	–	–	0.621	0.612	–	–	–	0.315	0.315
	$\rho\times\beta$	7.020	2.944	3.609	–	–	2.862	1.288	1.590	–	–	1.438	0.606	0.751	–	–	0.828	0.318	0.398	–	–
	ϕ	3.074	0.876	1.058	0.949	0.920	2.447	0.400	0.485	0.416	0.406	2.152	0.197	0.237	0.200	0.197	2.000	0.098	0.124	0.099	0.098
	π^+	–	–	–	16.431	15.864	–	–	–	6.759	6.628	–	–	–	3.043	3.013	–	–	–	1.432	1.421
	π^-	–	–	–	16.871	16.539	–	–	–	6.638	6.476	–	–	–	3.027	3.002	–	–	–	1.427	1.419
	$<\pi$	5.704	4.097	4.949	–	–	2.500	1.799	2.195	–	–	1.200	0.854	1.058	–	–	0.565	0.415	0.496	–	–

Note. Evaluation of the models' performance when DGP is linear, the case (iv) and $\theta = 5$. 50, 100, 200, and 400 indicate the sample sizes. ML represents the Johansen model, while ARDL is the auto-regressive distributed lag model. QARDL is the quantile ARDL, NARDL is the single-step nonlinear ARDL, and 2NARDL is the two–step NARDL.

Table 6. Monte Carlo simulation results: finite sample mean squared error (MSE) of the estimators (MSE × 100).

		50					100					200					400				
ϕ		ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL	ML	ARDL	QARDL	NARDL	2NARDL
−0.5	ρ	34.298	32.705	32.547	3.240	2.474	41.502	40.156	40.127	1.798	1.457	43.656	43.101	43.083	1.296	1.157	44.244	44.076	44.075	1.113	1.071
	$\rho \times \beta^+$	–	–	–	10.252	10.390	–	–	–	6.426	6.608	–	–	–	4.905	4.993	–	–	–	4.341	4.475
	$\rho \times \beta^-$	–	–	–	3.012	3.160	–	–	–	1.571	1.882	–	–	–	1.187	1.311	–	–	–	1.064	1.142
	$\rho \times \beta$	58.265	53.111	52.991	–	–	77.190	72.954	72.873	–	–	89.208	86.215	86.142	–	–	95.076	93.404	93.370	–	–
	ϕ	4.394	2.694	3.281	1.826	2.367	4.576	1.786	2.194	1.036	1.328	4.575	1.206	1.457	0.680	0.798	4.392	0.810	0.983	0.521	0.569
	π^+	–	–	–	21.119	21.127	–	–	–	8.929	8.823	–	–	–	4.100	4.077	–	–	–	2.028	2.026
	π^-	–	–	–	21.953	22.279	–	–	–	8.752	8.692	–	–	–	4.161	4.130	–	–	–	2.007	1.995
	π	9.964	18.305	20.627	–	–	4.171	18.242	19.287	–	–	1.875	18.164	18.757	–	–	0.896	18.113	18.220	–	–
0	ρ	34.773	32.083	32.293	2.255	2.157	42.049	40.567	40.626	1.354	1.384	43.976	43.428	43.435	1.124	1.162	44.343	44.197	44.203	1.087	1.112
	$\rho \times \beta^+$	–	–	–	8.494	8.698	–	–	–	5.335	5.692	–	–	–	4.474	4.729	–	–	–	4.347	4.492
	$\rho \times \beta^-$	–	–	–	2.721	2.377	–	–	–	1.411	1.477	–	–	–	1.133	1.195	–	–	–	1.091	1.126
	$\rho \times \beta$	70.997	63.454	64.166	–	–	86.977	83.748	83.906	–	–	95.114	93.322	93.375	–	–	98.114	97.233	97.234	–	–
	ϕ	8.676	1.696	2.336	1.728	1.698	9.059	0.763	1.073	1.022	1.056	8.939	0.397	0.541	0.727	0.758	8.710	0.227	0.304	0.597	0.612
	π^+	–	–	–	17.778	16.982	–	–	–	7.300	7.112	–	–	–	3.279	3.242	–	–	–	1.567	1.561
	π^-	–	–	–	18.212	17.650	–	–	–	7.099	6.871	–	–	–	3.266	3.230	–	–	–	1.535	1.523
	π	9.200	12.441	14.372	–	–	3.743	10.252	11.025	–	–	1.659	9.186	9.387	–	–	0.772	8.766	8.746	–	–
0.5	ρ	30.295	27.200	27.289	1.311	1.228	40.107	38.562	38.616	0.557	0.545	43.558	42.909	42.910	0.259	0.262	44.288	44.078	44.082	0.134	0.138
	$\rho \times \beta^+$	–	–	–	6.140	5.661	–	–	–	2.370	2.317	–	–	–	1.063	1.058	–	–	–	0.543	0.546
	$\rho \times \beta^-$	–	–	–	2.358	1.805	–	–	–	0.710	0.652	–	–	–	0.289	0.277	–	–	–	0.141	0.139
	$\rho \times \beta$	78.441	64.532	65.073	–	–	96.307	87.284	87.619	–	–	101.863	96.560	96.563	–	–	101.379	99.258	99.255	–	–
	ϕ	16.666	2.924	3.271	0.929	0.903	19.452	2.835	2.951	0.409	0.400	20.359	2.798	2.781	0.195	0.193	20.462	2.717	2.614	0.097	0.095
	π^+	–	–	–	16.398	15.876	–	–	–	6.787	6.663	–	–	–	3.064	3.036	–	–	–	1.443	1.433
	π^-	–	–	–	16.793	16.434	–	–	–	6.602	6.437	–	–	–	3.005	2.979	–	–	–	1.416	1.407
	π	11.861	13.019	15.595	–	–	5.020	10.218	11.389	–	–	2.202	8.788	9.141	–	–	1.006	8.295	8.171	–	–

Note. Evaluation of the models' performance when DGP is nonlinear, the case (iv) and $\theta = 5$. 50, 100, 200, and 400 indicate the sample sizes. ML represents the Johansen model, while ARDL is the auto-regressive distributed lag model. QARDL is the quantile ARDL, NARDL is the single-step nonlinear ARDL, and 2NARDL is the two-step NARDL.

Table 7. Estimates of the dynamic model of the unemployment–output relationship.

	ML		ARDL		QARDL		NARDL		2NARDL	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
UR_{t-1}	−0.0105	0.0000	−0.0233	0.0112	−0.0091	0.0179	−0.0437	0.0161	−0.0403	0.0112
IP_{t-1}	−0.0679	0.0003	−0.0399	0.0385	−0.1019	0.0578	–	–	–	–
IP^+_{t-1}	–	–	–	–	–	–	−0.0069	0.1616	−0.0726	0.0008
IP^-_{t-1}	–	–	–	–	–	–	−0.1472	0.4947	−0.0457	0.0005
ΔUR_{t-1}	0.2981	0.0001	0.2981	0.0602	0.2813	0.0851	0.2796	0.0577	0.2818	0.0569
ΔIP_t	−11.6392	0.0008	−11.6392	0.7831	−11.1781	1.3961	–	–	–	–
ΔIP^+_t	–	–	–	–	–	–	−7.1564	1.1826	−6.8993	1.1336
ΔIP^-_t	–	–	–	–	–	–	−17.0213	1.3693	−17.1542	1.3177
ΔIP_{t-1}	−2.7793	0.0012	−2.7793	1.1720	−3.5948	1.5012	–	–	–	–
ΔIP^+_{t-1}	–	–	–	–	–	–	−2.3363	1.3128	−2.1634	1.2659
ΔIP^-_{t-1}	–	–	–	–	–	–	−5.0349	1.8031	−5.3126	1.7600

Note. The table reports parameter estimates of the ARDL (2,2) model in the error correction form. ML represents the Johansen model, while ARDL is the auto-regressive distributed lag model. QARDL is the quantile ARDL, NARDL is the single-step non-linear ARDL, and 2NARDL is the two-step NARDL.

statistical tests. In our study, we adopt the same dataset as Balke and Fomby (1994) and Dijk et al. (1999), covering the period 1948Q1–1993Q4, and focus on two key variables: industrial production (IP), used as a proxy for economic growth, and the unemployment rate (UR).[6] Both series have been shown to contain additive and innovation outliers in this period. Rather than conducting new outlier detection procedures, we rely on the established findings of Balke and Fomby (1994), as our primary objective is not to detect outliers but to examine how a model can accommodate them. We assume that both variables have unit root, which is already expected for IP, we motivated by hypothesis of hysteresis for unemployment rate (see Blanchard and Summers 1986 for discussion).

We fit a second-order ARDL model to the data, which is supported by residual analysis. The data also fit a second-order NARDL model that allows for short- and long-run asymmetry. Table 7 presents the parameter estimates in error correction form. In the table, the standard errors of IP^+_{t-1} and IP^-_{t-1} for the two-step NARDL model are computed using the Delta method, while the standard errors for the Johansen model are computed *via* a bootstrap approach. According to the results, the long-run disequilibrium errors of the single- and two-step NARDL models are nearly identical. Both estimates indicate that disequilibria are corrected at a rate of approximately 4% per quarter. Aside from the long-run parameters, the differences in parameter estimates are minor; however, the two-step NARDL approach yields more precise estimates. Unlike the NARDL-based models, the speed of error correction in the linear models differs. The adjustment rates are 2.3% and 1% for the ARDL and ML models, respectively. The adjustment coefficient of the QARDL model is notably lower, at 0.9%. Considering the short-run coefficients, the ARDL and Johansen models have similar parameter estimates; nevertheless, the ML model produces much smaller standard errors. It should be noted that bootstrap standard errors are used for the ML model when interpreting the results. As the standard errors are derived using different methods across models, they are not directly comparable. Accordingly, our emphasis is placed on the parameter estimates rather than their associated standard errors.

[6]The data is taken from Journal of Applied Econometrics data archive http://econ.queensu.ca/jae/. Following Balke and Fomby (1994) and Dijk et al. (1999) logarithmic form of industrial production is used and seasonally adjusted data are employed for the both variables. Balke and Fomby (1994) identify several additive and innovation outliers in the post-WWII U.S. macroeconomic series we use. Specifically, they document additive outliers in the industrial production series in 1950Q3, 1958Q3, 1960Q1, 1975Q1, and 1978Q2, and innovation outliers in 1959Q3. For the unemployment rate, they detect a large additive outlier in 1954Q1 and an innovation outlier in 1975Q1. These outliers are closely linked to major macroeconomic disruptions, including recessions and structural shifts. In this study, we rely on their identification of these outliers to motivate our use of models robust to such shocks.

The empirical and simulation results are broadly consistent. Both the single- and two-step nonlinear ARDL models perform similarly, with the two-step version generally yielding better results in all cases. For the linear models, simulations show that the ARDL outperforms the QARDL in almost all scenarios and is also more efficient than the Johansen model. The empirical application indicates that the Johansen and ARDL models yield closely similar parameter estimates.

5. Concluding remarks

Cointegration models are introduced under the assumption of normality, but in empirical applications, this assumption is frequently violated by unexpected events, such as policy implementations, external shocks or measurements errors. Such events usually show up as abnormal observations and structural breaks in data. As shown in the Monte Carlo simulation parts of this paper, estimators of cointegration models are sensitive to those types of data irregularities.

In this paper, we focus on widely used cointegration models and investigate their performance in the presence of additive outliers. Our findings indicate that the performance of the models depends the magnitude and type of outliers as well as the sample size. Overall, according to the main results, the Johansen model is the most vulnerable among linear models when data contains AOs while the ARDL model is the less sensitive one. The results are consistent with Franses and Haldrup (1994) and Franses and Lucas (1998), which find the Johansen ML estimator is less efficient in the presence of AOs. On the other hand, although there are some exceptions, single- and two-step nonlinear ARDL estimations yield similar results, especially if the sample size is large. The result is supported by Cho et al. (2019), these findings are also supported by the empirical example. The empirical analysis results indicate that two-step NARDL generates more precise estimates than single-step NARDL. The Johansen and ARDL models have the same parameters but the former yields more accurate estimates than the latter.

An inherent challenge in empirical macroeconomic analysis is the potential presence of structural breaks, particularly when working with long time series. These breaks may arise from major policy changes, economic crises, or shifts in institutional frameworks, each of which can affect the stability of estimated relationships. While using longer samples enhances statistical power and allows more robust inference in simulation studies, it also increases the likelihood of capturing multiple structural regimes, potentially biasing the results or complicating their interpretation. We take this limitation into account when interpreting the findings.

Outliers frequently arise in empirical applications, but they do not always represent "bad" or erroneous observations. In many cases, they can provide valuable information about model misspecification, structural breaks, or omitted variables, and thus highlight directions for model refinement (see Franses et al. 1998; Franses and Lucas 1998). Nevertheless, it is well known that outliers can distort statistical inference and cause misleading conclusions, especially in small samples. To mitigate such risks, applied researchers often resort to dummy-variable adjustments, although Nielsen (2004) cautions that misspecified dummies can be more detrimental than ignoring the outliers altogether. Alternatively, robust estimation techniques, such as maximum likelihood-type (M) estimators (see Huber 1981), offer formal ways to reduce the influence of atypical observations.

The situation becomes even more complex in multivariate models, where outliers may affect different variables in different time periods. Exploring all possible configurations is often computationally infeasible. While our study does not incorporate robust estimation procedures such as those developed by Franses and Lucas (1998), we acknowledge the importance of this line of research. Our focus is instead on empirically comparing the performance of widely used cointegration estimators in the presence of additive outliers. Future work could extend our findings by integrating generalized pseudo-likelihood or robust M-estimators into ARDL-based frameworks, and evaluating their performance under contamination. Such extensions would provide both theoretical and practical insights into improving inference in the presence of outliers.

Acknowledgements

The authors thank the editor and reviewers for their valuable comments and suggestions for improvement.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Open access funding provided by Uppsala University.

References

Ankargren S, Lyhagen J. 2019. Estimating a VECM for a small open economy. Working Paper 2019: 2. Department of Statistics, Uppsala University.
Balke NS, Fomby TB. 1994. Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series. J Appl Econ. 9(2):181–200. https://doi.org/10.1002/jae.3950090205
Blanchard OJ, Summers LH. 1986. Hysteresis and the European unemployment problem, Working Paper 1950, National Bureau of Economic Research.
Cho JS, Greenwood-Nimmo M, Shin Y. 2019. Two-step estimation of the nonlinear autoregressive distributed lag model, Working Papers 2019rwp-154, Yonsei University, Yonsei Economics Research Institute.
Cho JS, Kim T, Shin Y. 2015. Quantile cointegration in the autoregressive distributed-lag modeling framework. J Econ. 188(1):281–300. https://doi.org/10.1016/j.jeconom.2015.05.003
Desgagné A. 2021. Efficient and robust estimation of regression and scale parameters, with outlier detection. Comput Stat Data Anal. 155:107114. https://doi.org/10.1016/j.csda.2020.107114
Dijk DV, Franses PH, Lucas A. 1999. Testing for arch in the presence of additive outliers. J Appl Econ. 14:539–562.
Doornik JA, Ooms M. 2005. Outlier detection in GARCH models. Tinbergen Institute Discussion Papers 05-092/4, Tinbergen Institute.
Dunsch S. 2017. Age- and gender-specific unemployment and Okun's law in CEE countries. East Eur Econ. 55(4): 377–393. https://doi.org/10.1080/00128775.2017.1338962
Fox AJ. 1972. Outliers in time series. J Roy Stat Soc Ser B (Method). 34(3):350–363. https://doi.org/10.1111/j.2517-6161.1972.tb00912.x
Franses PH, Ghijsels H. 1999. Additive outliers, GARCH and forecasting volatility. Int J Forecast. 15(1):1–9. https://doi.org/10.1016/S0169-2070(98)00053-3
Franses PH, Haldrup N. 1994. The effects of additive outliers on tests for unit roots and cointegration. J Business Econ Stat. 12(4):471–478. https://doi.org/10.1080/07350015.1994.10524569
Franses PH, Kloek T, Lucas A. 1998. Outlier robust analysis of long-run marketing effects for weekly scanning data. J Econ. 89(1–2):293–315. https://doi.org/10.1016/S0304-4076(98)00065-7
Franses PH, Lucas A. 1998. Outlier detection in cointegration analysis. J Business Econ Stat. 16(4):459–468. https://doi.org/10.1080/07350015.1998.10524785
Granger C, Newbold P. 1974. Spurious regressions in econometrics. J Econ. 2(2):111–120. https://doi.org/10.1016/0304-4076(74)90034-7
Granger CWJ, Yoon G. 2002. Hidden cointegration, Working Paper 2002-02. University of California.
Huber PJ. 1981. Robust statistics. John Wiley & Sons.
Johansen S. 1995. Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press.
Koenker R, Bassett G. 1978. Regression quantiles. Econometrica. 46(1):33–50. https://doi.org/10.2307/1913643
Kreishan FM. 2011. Economic growth and unemployment: an empirical analysis. Journal of Social Sciences, 2, 7, 228–231. https://doi.org/10.3844/jssp.2011.228.231
Lang D, de Peretti C. 2009. A strong hysteretic model of Okun's law: theory and a preliminary investigation. Int Rev Appl Econ. 23(4):445–462. https://doi.org/10.1080/02692170902954775

Lee J. 2000. The robustness of Okun's law: evidence from OECD countries. J Macroecon. 22(2):331–356. https://doi.org/10.1016/S0164-0704(00)00135-X

Lee K et al. 2022. Outliers in financial time series data: outliers, margin debt, and economic recession. Mach Learn Appl. 10:100420. https://doi.org/10.1016/j.mlwa.2022.100420

Loperfido N. 2020. Kurtosis-based projection pursuit for outlier detection in financial time series. Eur J Finance. 26(2–3):142–164. https://doi.org/10.1080/1351847X.2019.1647864

Lucas A. 1997. Cointegration testing using pseudolikelihood ratio tests. Econ Theory. 13(2):149–169. https://doi.org/10.1017/S0266466600005703

Maddala G, Kim I. 1998. Unit roots, cointegration, and structural change, themes in modern econometrics. Cambridge University Press.

Nielsen HB. 2004. Cointegration analysis in the presence of outliers. Econ J. 7:249–271.

Okun A. 1963. Potential GNP: its measurement and significance. Cowles Foundation Paper, Cowles Foundation for Research in Economics at Yale University.

Pesaran MH, Shin Y. 1999. An autoregressive distributed-lag modelling approach to cointegration analysis. Cambridge University Press. p. 371–413.

Pesaran MH, Shin Y, Smith RJ. 2001. Bounds testing approaches to the analysis of level relationships. J Appl Econ. 16(3):289–326. https://doi.org/10.1002/jae.616

Phillips P, Hansen B. 1990. Statistical inference in instrumental variables regression with I(1) processes. Rev Econ Stud. 57(1):99–125. https://doi.org/10.2307/2297545

Premaratne G, Bera AK. 2000. Modeling asymmetry and excess kurtosis in stock return data. Capital markets: asset pricing and valuation eJournal.

Schorderet Y. 2001. Revisiting Okun's law: an hysteretic perspective. Working Paper 2001-13, University of California at San Diego, Economics Working Paper Series qt2fb7n2wd, Department of Economics. UC San Diego.

Schorderet Y. 2003. Asymmetric cointegration. Working Paper 2003-1. University of Geneva.

Shahzad SJH, Hurley D, Ferrer R. 2021. U.S. stock prices and macroeconomic fundamentals: fresh evidence using the quantile ARDL approach. Int J Fin Econ. 26(3):3569–3587. https://doi.org/10.1002/ijfe.1976

Shin Y, Yu B, Greenwood-Nimmo M. 2014. Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework. Springer. p. 281–314.

Sögner L, Stiassny A. 2002. An analysis on the structural stability of Okun's law – a cross-country study. Appl Econ. 34:1775–1787.

Tolvi J. 2001. Outliers in eleven Finnish macroeconomic time series. Finn Econ Pap. 14:14–32.

Wagner M, Hong SH. 2016. Cointegrating polynomial regressions: fully modified OLS estimation and inference. Econ Theory. 32(5):1289–1315. https://doi.org/10.1017/S0266466615000213