ISSN: 2836-8495

Research Article

Journal of Current Trends in Computer Science Research

ARDL Modeling Using R Software

Sami Mestiri*

University of Monastir, Tunisia. Rue Ibn Sina Hiboun, Mahdia Tunisia.

***Corresponding Author**
Sami Mestiri, University of Monastir, Tunisia. Rue Ibn Sina Hiboun, Mahdia Tunisia.

Submitted: 2023, Dec 12; **Accepted:** 2024, Jan 09; **Published:** 2024, Feb 13

Citation: Mestiri, S. (2024). ARDL Modeling Using R Software. *J Curr Trends Comp Sci Res, 3*(1), 01-05.

Abstract

The goal of this paper is helping to apply ARDL models using the R software. We will cover its benefits, show how to use the packages and will make interesting recommendations for estimating models ARDL using R.

This paper presents the dynamac package for the statistical language R, demonstrating its main functionalities in a step by step guide.

JEL codes: C15, C88

Keywords: R Software, ARDL, Cointegration Test.

1. Introduction

Pesaran et al. (2001) introduced the bounds test for cointegration based on the previous work of Pesaran and Shin (1999) using the ARDL model as a platform for the test. Since then, the ARDL framework and the bounds test are used constantly by practitioners who seem to adopt every new advancement of the initial framework. A recent example combining various techniques, is Wu et al. (2022) who applied bootstrap ARDL with a Fourier function. This paper provides a smooth introduction to the dynamac package in R and its main features and capabilities.

Regarding proprietary software like EViews, although they are generally considered more user-friendly, they lack flexibility compared to programming languages such as R. Additionally, these software platforms are often slow to adopt the latest advancements in research and can be prohibitively expensive for many users.

On the other hand, open-source software does not provide any guarantees regarding the quality of results, and it is the responsibility of the user to verify the code. The problem lies in the fact that not everyone is an expert in the field, making it challenging to technically validate the code's implementation. Many practitioners simply seek reliable software they can trust.

Dynamac is a suite of programs in R designed to assist users in modeling and visualizing the effects of autoregressive distributed lag models, as well as testing for cointegration. The core program is dynardl, a flexible program designed to dynamically simulate and plot a variety of types of autoregressive distributed lag models, including error-correction models.

The research paper is organized as follows: We provide Auto Regressive Distributed Lag models in Section 2. Section 3 presents Cointegration test. In section 4, we apply the model. And finally, we conclude in section 5.

2. Auto Regressive Distributed Lag models

Auto Regressive Distributed Lag models (ARDL), are dynamic models which involve variables lagged over time unlike static models. These models have the particularity of considering temporal dynamics (adjustment time, expectations, etc.) in the explanation of a variable (time series), thus improving the forecasts and effectiveness of policies (decisions, actions, etc.), unlike the simple (nondynamic) model whose instantaneous explanation (immediate effect or not spread over time) only restores part of the variation in the variable to explain.

In ARDL models we find, among the explanatory variables (X_t), the lagged dependent variable (Y_{t-p}) and the past values of the independent variable (X_{t-q}). They have the following general form::

$$Y_t = f(X_t, Y_{t-p}, X_{t-q})$$

In its general (explicit) form, an ARDL model is written as follows:

$$Y_t = a_0 + a_1 Y_{t-1} + \cdots + a_p Y_{t-p} + b_0 X_t + + \cdots + b_q X_{t-q} + \varepsilon_t$$

$$Y_t = a_0 + \sum_{i=1}^{p} Y_{t-i} + \sum_{j=0}^{q} X_{t-j} + \varepsilon_t$$

With $\varepsilon \sim (0,\sigma)$ error term.
b_0 translates the short-term effect of X_t on Y_t.
If we consider the following long-term or equilibrium relationship $Y_t = k + \varphi X_t + u$, we can calculate the long-run effect of X_t on Y_t as follows:

$$\phi = \frac{\sum b_j}{1 - \sum a_i}$$

As with any dynamic model, we will use the information criteria (AIC, SIC and HQ) to determine the optimal shift (p* or q*); an optimal shift is one whose estimated model offers the minimum value of one of the stated criteria. These criteria are: that of Akaike (AIC), that of Schwarz (SIC) and that of Hannan and Quinn (HQ). Their Akaike values (AIC) are calculated as follows:

$$AIC\,(h) = Ln\left(\frac{SCR_h}{n}\right) + \frac{2h}{n}$$

with SCR_h = Sums of Squares of Residuals for the model with h delays
n = Number of observations
Ln = Natural logarithm

These ARDL models generally suffer from error autocorrelation problems, with the presence of the lagged endogenous variable as explanatory and from multi-collinearity, which complicates the estimation of the parameters by Ordinary Least Squares/ OLS. Here, he has to resort to techniques robust estimation (SUR method, etc.) to overcome these problems. Also, we note that the variables considered in these models must be stationary to avoid spurious regressions. The ARDL model makes it possible to estimate short-term dynamics and long-term effects for cointegrated series or even integrated at different orders.

3. Cointegration Test

When we have several integrated variables of different orders (I(0), I(1)), we can use the cointegration test of Pesaran et al. (2001) called "bounds test to cointegration", initially developed by Pesaran and Shin (1999).

The model which serves as a basis for the test of cointegration by staggered lags (test of Pesaran et al. (2001)) is the following cointegrated ARDL specification (it takes the form of an error correction model or a VECM), when we study the dynamics between two series x_t and y_t.

$$\Delta y_t = \gamma_1 y_{t-1} + \gamma_2 x_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \sum_{j=0}^{q-1} \beta_j \Delta x_{t-j} + \pi_0 + \pi_t + e_t$$

This specification presents the ARDL model which can be written as follows:

$$\Delta y_t = \pi_0 + \pi_t + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \sum_{j=0}^{q-1} \beta_j \Delta x_{t-j} + \lambda \varepsilon_{t-1} + e_t$$

Whereù λ is the error correction term, adjustment coefficient or restoring force. we conclude à the existence of a cointegration relation between x_t and y_t if and only if $0 \prec \left|\widehat{\lambda}\right| \prec 1$ and rejection $H_0 : \lambda = 0$.

There are two à steps to follow to apply the Pesaran cointegration test, namely: the determination of the eoptimal calibration above all (AIC, SIC) and uses the Fisher test to verify the hypotheses:
$H_0 : \alpha_1 = \alpha_2 = 0$ existence of a cointegration relation
$H_1 : \alpha_1$ 6= α_2 6= 0 absence of a cointegration relation

The test procedure is such that we must compare the Fisher values obtained with the critical values (bounds) simulated for several cases and different thresholds by Pesaran et al. We will note from the critical values that the upper bound takes up the values for which the variables are integrated of order 1 I (1)and the lower bound concern the variables I (0).

Thus: $F_c \succ B^{\sup}$ Cointegration exists
$F_c \prec B^{\inf}$ Cointegration does not exist
$B^{\inf} \prec F_c \prec B^{\sup}$ There is no conclusion

4. Application

We illustrate the process autoregressive distributed lag modeling, testing for cointegration with pssbounds, and interpretation of X through stochastic simulations using data originally from Wright (2017) on public concern about inequality in the United States.7 For our example, assume that public concern about inequality in the US, Concern (concern), is a function of the share of income going to the top ten percent, Income Top 10 (incshare10). We also hypothesize that the unemployment rate, Unemployment (urate), affects concern over the short-run (i.e., is not cointegrating): Before estimating any model using dynamac, users should first check for stationarity. A variety of unit root tests can be performed using the urca package [1]. These suggest that all three series are integrated of order I(1), as they appear integrated in levels but stationary in first-differences (D), shown in Table 1

	Augmented DF	Phillips-Perron	Dickey-Fuller GLS	KPSS
Concern	0.688	-3.437*	-0.893	0.642*
ΔConcern	-3.507*	-7.675*	-3.124*	0.814*
Unemployment	-0.612	-2.762	-2.802*	0.224
ΔUnemployment	-5.362*	-4.879*	-5.308*	0.064
Income Top 10	2.992	0.442	0.994	2.482*
ΔIncome Top 10	-3.170*	-6.244*	-4.032*	0.218

One augmenting lag included for all tests. $*: p < 0.05$. Augmented Dickey-Fuller, PP, and DF-GLS have null hypothesis of a unit-root, while KPSS has a null of stationarity.

Table 1: Unit Root Tests

Given that all series appear to be I(1), we proceed with estimating a model in dynardl in error correction form, and then testing for cointegration between concern about inequality and the share of income of the top 10 percent. In general, we suggest using this strategy outlined in Philips (2018) along with alternative tests for cointegration. Our error-correction model appears as:

$$\Delta Concernt = a0 + \phi_1 Concern_{t-1} + \Delta_1 IncomeTop10_{t-1} + \beta_1 \Delta IncomeTop10_t + \beta_2 \Delta Unemployment_t + \epsilon_t$$

where we assume $\epsilon_t \sim N(0, \sigma^2)$.

dynardl is simply an engine for regression, but one that allows users to focus on theoretical specification rather than technical coding. All variables in the model are entered into the formula. In this sense, dynardl can be used in any ARDL context, not just ones in which the user is also expecting cointegration testing or dynamic simulations.We estimate our example model shown in Equation using dynardl as follows: data(ineq) res1 <- dynardl(concern ~ incshare10 + urate, data = ineq, lags = list("concern" = 1, "incshare10" = 1), diffs = c("incshare10", "urate"),
ec = TRUE, simulate = FALSE)summary(res1)

As shown from the regression results, dynardl has included a constant, the lagged dependent variable, l.1.concern, the first difference of the two regressors (Income Top 10 and Unemployment), as well as the lag of Income Top 10.

While changes in Income Top 10 affect changes in Concern in the short-run, changes in Unemployment do not have a statistically significant effect in the short-run. The lag of Income Top 10 is negative and statistically significant.

```
Call:
lm(formula = as.formula(paste(paste(dvnamelist), "~", paste(colnames(IVs)
    collapse = "+"), collapse = " ")))

Residuals:
      Min        1Q    Median        3Q       Max
-0.025848 -0.005293  0.000692  0.006589  0.031563

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.122043   0.027967   4.364 7.87e-05 ***
l.1.concern    -0.167655   0.048701  -3.443   0.0013 **
d.1.incshare10  0.800585   0.296620   2.699   0.0099 **
d.1.urate       0.001118   0.001699   0.658   0.5138
l.1.incshare10 -0.068028   0.031834  -2.137   0.0383 *
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01169 on 43 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.3671,       Adjusted R-squared:  0.3083
F-statistic: 6.236 on 4 and 43 DF,  p-value: 0.0004755
```

Table 2:

As shown from the regression results, dynardl has included a constant, the lagged dependent variable, l.1.concern, the first difference of the two regressors (Income Top 10 and Unemployment), as well as the lag of Income Top 10.

While changes in Income Top 10 affect changes in Concern in the short-run, changes in Unemployment do not have a statistically significant effect in the short-run. The lag of Income Top 10 is negative and statistically significant.

Moreover, the parameter on the lagged dependent variable is negative, between 0 and -1, and statistically significant, giving us cursory evidence of a cointegrating process taking place; we use a statistical test for this below.

An essential component of ARDL modeling is ensuring that the residuals from any ARDL estimation are white noise. One symptom of residual autocorrelation in the presence of a lagged dependent variable (where f1 6= 0) is that OLS will result in biased and inconsistent estimates. Autocorrelation is especially pernicious when using the ARDLbounds cointegration test, since the test relies on the assumption of, serially uncorrelated errors for the validity of the bound's tests.

To assist users in model selection and residual testing, we offer dynardl.auto.correlated. This function takes the residuals from an ARDL model estimated by dynardl and conducts two tests for autocorrelation the Shaprio-Wilk test for normality and the Breusch-Godfrey test for higher-order serial correlation as well as calculates the fit statistics for the Akaike information criterion (AIC), Bayesian information criterion (BIC), and log-likelihood.

```
res2 <- dynardl(concern ~ incshare10 + urate, data = ineq,
        lags = list("concern" = 1, "incshare10" = 1),
        diffs = c("incshare10", "urate"),
        lagdiffs = list("concern" = 1),
        ec = TRUE, simulate = FALSE)

pssbounds(res2)

------------------------------------------------------
-                       F-test                       -
------------------------------------------------------
                <------- I(0) ------------ I(1) ----->
10% critical value       4.19             4.94
5% critical value        5.22             6.07
1% critical value        7.56             8.685


F-statistic = 12.204
------------------------------------------------------
```

Table 3:

In our example, since the value of the F-statistic exceeds the critical value at the upper I(1) bound of the test at the 1% level, we may conclude that Income Top 10 and Concern about inequality are in a cointegrating relationship. As an auxiliary test, pssbounds displays a one-sided test on the t-statistic on the lagged dependent variable.

Since the t-statistic of -3.684 falls below the 5% critical value I(1) threshold, this lends further support for cointegration. Taken together, these findings indicate that there is a cointegrating relationship between concern about inequality and the income share of the top 10 percent, and that Equation 6 is appropriately specified.

5. Conclusions

This paper serves as a comprehensive step-by-step guide, showcasing the core functionalities of the dynamac package, a versatile tool developed in the R language. In addition to explaining the package capabilities, we provide simple examples that end-users can readily adopt and tailor to suit their unique research requirements.

Throughout the illustrative examples, we highlight the user-friendly dynamac package, which enables effortless estimation of even the most intricate models. The package flexibility becomes evident as it easily accommodates the calculation of complex designs, making it a valuable asset for researchers seeking reliable and robust results [2-10].

References

1. Pfaff, B., Zivot, E., Stigler, M., & Pfaff, M. B. (2016). Package 'urca'. *Unit root and cointegration tests for time series data. R package version*, 1-2.
2. Jordan, S., Philips, A. Q. (2022). Dynamac: Dynamic simulation and testing for single-equation ARDL models.
3. Mestiri, S. (2020). Using R software to applied econometrics
4. Mestiri, S. (2019) How to use the R software. University of Monastir Press. DOI 10.13140/RG.2.2.18152.83206
5. Mestiri, S., & Farhat, A. (2021). Using non-parametric count model for credit scoring. *Journal of Quantitative Economics*, 19, 39-49.
6. Mestiri, S. (2021). Bayesian Structural Var Approch To Tunisian Monetary Policy Farmework. *Journal of Smart Economic Growth*, 6(2), 67-77.
7. Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of applied econometrics,* 16(3), 289-326.
8. Pesaran, M. H., & Shin, Y. (1995). *An autoregressive distributed lag modelling approach to cointegration analysis* (Vol. 9514). Cambridge, UK: Department of Applied Economics, University of Cambridge.
9. Zeileis, A., & Zeileis, M. A. (2019). Package 'dynlm'.
10. Wright, G. (2018). The political implications of American concerns about economic inequality. *Political Behavior,* 40, 321-343.