# ARDL: An R package for the analysis of level relationships

Kleanthis Natsiopoulos [1] and **Nickolaos G. Tzeremes** [1]

1 Department of Economics, University of Thessaly, 28th October, 78, 38333, Volos, Greece

Summary

Autoregressive Distributed Lag (ARDL) and Error Correction Models (ECM) (Pesaran & Shin, 1999) are widely used in various economic, environmental, political science applications etc. These are very flexible because of the autoregressive (AR) and the distributed lag (DL, essentially AR terms of the independent variables) terms in the ARDL model. Also, they are used in the context of cointegration analysis as a platform to test and analyze the levels (long-run) relationship between variables. One of the most popular such tests is the bounds test proposed by Pesaran et al. (2001) which allows testing for cointegration while at the same time estimates the level relationship. ARDL (Natsiopoulos & Tzeremes, 2021; Natsiopoulos & Tzeremes, 2022) is an R package that aims to help users in the modeling process of ARDL and ECM and it also provides the tools towards the bounds test for cointegration.

Statement of need

ARDL is implemented in such a way that researchers can use it as a full featured tool for this specific type of analysis and students of all levels can be aware of how each piece of code works through the analytical manual and the examples which cover every functionality of the package. One of the greatest advantages of the package is that there is no need for the user to know the complex mathematical equations and write them by hand. Instead, the correct specification of the model is created automatically, just by providing the order of the ARDL. Some recent examples of publications that used the package are Baruah (2022) and Qiu et al. (2021).

It is very important to note that the ARDL package is the first software which managed to fully replicate the original paper by Pesaran et al. (2001), verifying the validity and accuracy of the package (Natsiopoulos & Tzeremes, 2022).

State of the field

ARDL distinguishes itself from other related R packages like dLagM (Demirhan, 2020) and dynamac (Jordan & Philips, 2020) in the sense that it is specifically designed to address a particular problem throughout every phase of modeling, testing and interpretation. It includes a rich set of dedicated tools, accompanied by an analytical manual that describes the mathematical process of every function.

On the other hand, dLagM and dynamac provide additional plotting functionalities and post-estimation diagnostics (serial correlation tests etc). The design of the ARDL packages differs from the pre-mentioned packages as it is not an *all-in-one* package. It is designed so that every exported object is of commonly used R classes so that it can be easily combined with

other packages, each of which are also dedicated to a specific part of the post-estimation part of the analysis.

In addition, the `ARDL` packages natively supports time-series data, thus sub-sample and balanced sample estimations are possible. It also allows to specify any level of significance for the bounds F-test and t-tests, and includes p-values and exact sample critical values for any possible combination. Also, the estimation of long-run, short-run and interim multipliers accompanied by standard errors and p-values as well as the cointegrating equation are some of the available features.

In the following example, we show how the estimation of an ARDL model can be performed using the `ARDL` package, as opposed to when we use a package specialized in dynamic time series modeling (i.e. `dynlm` (Zeileis, 2019) which is also the package that the `ARDL` package uses under the hood).

For this example we use the Danish data on money income prices and interest rates, from Johansen & Juselius (1990). We model the logarithm of real money, M2 (LRM) using the independent variables LRY, IBO and IDE (see ?denmark), and we estimate the ARDL(3,1,3,2).

Using the `ARDL` package we would write the following code:

```
library(ARDL)
```

```
ardl(LRM ~ LRY + IBO + IDE, data = denmark, order = c(3,1,3,2))
```

While using the `dynlm` package:

```
library(dynlm)
```

```
dynlm(LRM ~ L(LRM, 1) + L(LRM, 2) + L(LRM, 3) + LRY + L(LRY, 1) +
            IBO + L(IBO, 1) + L(IBO, 2) + L(IBO, 3) +
            IDE + L(IDE, 1) + L(IDE, 2), data = denmark)
```

Note that if someone wanted to estimate the same regression using the native `lm` function this would be even harder as the lags and differences are not supported inside the `lm` environment.

To convert this ARDL(3,1,3,2) model into an ECM, using the `ARDL` package we would write:

```
uecm_model <- uecm(ardl_model)
```

While using the `dynlm` package:

```
dynlm(d(LRM) ~ L(LRM, 1) + L(LRY, 1) + L(IBO, 1) +
                L(IDE, 1) + d(L(LRM, 1)) + d(L(LRM, 2)) +
                d(LRY) + d(IBO) + d(L(IBO, 1)) + d(L(IBO, 2)) +
                d(IDE) + d(L(IDE, 1)), data = denmark)
```

The API of `dynamac` has a similar approach with `dynlm` and one has to define every independent variable one by one. The API of `dLagM` requires less typing and the way for one to define an ARDL model is by setting the maximum lags for the dependent and the independent variables (the same maximum order for all the independent ones as in a typical VAR model) and then remove one by one the lags that are not required from each independent variable.

As this is a quick presentation of just one of the functionalities of the package, we are not going to list all the key advantages of the `ARDL` package or compare them with the limitations in other open source or proprietary software. We refere to these two aforementioned packages as these seem to be the R packages that argue that they support the same functionalities as the `ARDL` package. We believe that using the `ARDL` package is simpler, faster and less prone to errors.

Breaking down the ARDL package

Model estimation

The package does not explicitly connects the modeling with the bounds test, as the ARDL and ECM model may well be used independently in other research approaches. This way each function is dedicated to perform a very specific part of the whole process.

An $ARDL(p, q_1, ..., q_k)$ model can be fully specified by its order. The `ardl` function implements the following formula:

$$y_t = c_0 + c_1 t + \sum_{i=1}^{p} b_{y,i} y_{t-i} + \sum_{j=1}^{k} \sum_{l=0}^{q_j} b_{j,l} x_{j,t-l} + \epsilon_t \qquad (1)$$

An Unrestricted ECM (UECM) has a 1:1 relationship with the $ARDL(p, q_1, ..., q_k)$ model. The `uecm` function estimates it using the formula:

$$\Delta y_t = c_0 + c_1 t + \pi_y y_{t-1} + \sum_{j=1}^{k} \pi_j x_{j,t-1} + \sum_{i=1}^{p-1} \psi_{y,i} \Delta y_{t-i} + \sum_{j=1}^{k} \sum_{l=1}^{q_j-1} \psi_{j,l} \Delta x_{j,t-l} + \sum_{j=1}^{k} \omega_j \Delta x_{j,t} + \epsilon_t$$
$$(2)$$

The Restricted ECM (RECM) can also be fully described by the order of the underlying ARDL and the case for potential restriction in the deterministic parameters (constant and linear trend), using the function `recm`.

Model relationships

Figure 1 shows the relationships between the regression model objects. The bounds tests and all the other functions of the package also have this kind of interconnectivity as they inherit all the necessary information.
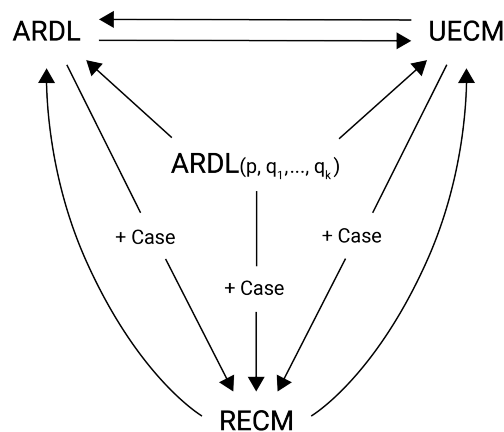

Figure 1: Interconnection between ARDL, UECM and RECM.

Natsiopoulos, & Tzeremes. (2022). ARDL: An R package for the analysis of level relationships. *Journal of Open Source Software*, 7(79), 3496. https://doi.org/10.21105/joss.03496.

Where $ARDL(p, q_1, ..., q_k)$ represent the ARDL order and Case is the restriction of the deterministic parameters. When an arrow points from one model or information (order or case) to another model, it can be interpreted as the first one can fully describe the second one. When it crosses the case it means that this information is also needed.

Bounds test

The functions `bounds_f_test` and `bounds_t_test` return a typical `htest` object and they perform a Wald or t test respectively on an UECM. The input of the functions can also be an ARDL model as they are interconnected as described above. The hypothesis tests for the bounds F-test and t-test are based on (2).

For the bounds F-test the null hypothesis is $\mathbf{H_0} : \pi_y = \pi_1 = \cdots = \pi_k = c_0 = c_1 = 0$, where the restriction of the deterministic trends c_0 and c_1 depends on the case chosen.

The null hypothesis of the bounds t-test is $\mathbf{H_0} : \pi_y = 0$.

Making inference after cointegration

After the modeling part and if a cointegrating relationship can be established, the long-run (but also the short-run or interim) multipliers can be computed using the `multipliers` function. Note that it is irrelevant whether they are estimated based on an ARDL (1) or an UECM (2), in terms of the results.

Lastly, the cointegrating relationship vector can be constructed using the function `coint_eq`.

Conclusion

Using the intuitive API of the `ARDL` package, even the most complex ARDL, UECM or RECM model can be fully specified using the $ARDL(p, q_1, ..., q_k)$ order and the case for the restriction of the constant and linear deterministic parameters. The interconnection between the objects allows for further post-estimation testing. Finally, the detailed estimation formulas are provided in order to enhance the understanding of every step of the analysis giving a reference point for the equations in publications.

Acknowledgments

We want to acknowledge Nikolaos Chatsios for designing the package's logo and the diagram in Figure 1.

References

Baruah, P. (2022). Investigating commodity price relations across wholesale markets: The case of paddy in chhattisgarh, india. *Indian Journal of Agricultural Economics*, *77*(1).

Demirhan, H. (2020). dLagM: An R package for distributed lag models and ARDL bounds testing. *PLoS ONE*, *15*(2), e0228812. https://doi.org/10.1371/journal.pone.0228812

Johansen, S., & Juselius, K. (1990). Maximum Likelihood Estimation and Inference on Cointegration – with Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics*, *52*(2), 169–210. https://doi.org/10.1111/j.1468-0084.1990.mp52002003.x

Jordan, S., & Philips, A. Q. (2020). *Dynamac: Dynamic simulation and testing for single-equation ARDL models*. https://CRAN.R-project.org/package=dynamac

Natsiopoulos, K., & Tzeremes, N. (2021). *ARDL: ARDL, ECM and bounds-test for cointegration*. https://CRAN.R-project.org/package=ARDL

Natsiopoulos, K., & Tzeremes, N. G. (2022). ARDL bounds test for cointegration: Replicating the pesaran et al. (2001) results for the UK earnings equation using r. *Journal of Applied Econometrics*, *37*(5), 1079–1090. https://doi.org/10.1002/jae.2919

Pesaran, M. H., & Shin, Y. (1999). An autoregressive distributed-lag modelling approach to cointegration analysis. In S. Strøm (Ed.), *Econometrics and economic theory in the 20th century: The ragnar frisch centennial symposium* (pp. 371–413). Cambridge University Press. https://doi.org/10.1017/CCOL521633230.011

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. https://doi.org/10.1002/jae.616

Qiu, R. T., Wu, D. C., Dropsy, V., Petit, S., Pratt, S., & Ohe, Y. (2021). Visitor arrivals forecasts amid COVID-19: A perspective from the asia and pacific team. *Annals of Tourism Research*, *88*, 103155. https://doi.org/10.1016/j.annals.2021.103155

Zeileis, A. (2019). *dynlm: Dynamic linear regression*. https://CRAN.R-project.org/package=dynlm