

Exploring meaningful visual effects and quantities of interest from dynamic models through dynamac

Soren Jordan¹ and Andrew Q. Philips²

¹ Assistant Professor, Department of Political Science, Auburn University ² Assistant Professor, Department of Political Science, University of Colorado Boulder

DOI: [10.21105/joss.02528](https://doi.org/10.21105/joss.02528)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Sebastian Benthall](#) ↗

Reviewers:

- [@stulacy](#)
- [@dormezil](#)

Submitted: 27 May 2020

Published: 05 October 2020

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

dynamac (2020) implements a cohesive ecosystem of estimating and interpreting autoregressive distributed lag (ARDL) models. Further, dynamac uses stochastic simulation techniques (Jordan & Philips, 2018a, 2018b) to easily recover traditional quantities of interest, such as short- and long-run effects, even from complex dynamic specifications (such as the addition of multiple lags of dependent variables, or independent variables appearing in multiple lags or multiple forms like contemporaneous and lagged differences, or contemporaneous and lagged levels), including models with cointegration (Philips, 2018). These simulation techniques bring a wider set of inferences from complex models to users in fields as diverse as environmental science (Danish, 2020; Liu & Liang, 2019; Sarkodie et al., 2020, 2019), epidemiology (Sarkodie & Owusu, 2020), economics (Khan, Teng, & Khan, 2019; Sharma, 2019), political science (Benton & Philips, 2020), and beyond. Users can make dynamic inferences from their models, including measures of uncertainty, without the need for any formulae or algebraic solutions. Although other R packages may estimate ARDL models (for instance dynlm (Zeileis, 2019), which only provides model estimates; ardl (Natsopoulos & Tzeremes, 2020), which uses dynlm for estimation but provides support for the bounds test for cointegration as well as automated lag selection; dlagM (Demirhan, 2020), which allows n-step model forecasting), dynamac is unique in that it is the only R package that provides post-estimation diagnostics for autocorrelation in the residuals and is designed specifically for providing inferences through counterfactual simulation.

The stochastic simulation procedure is as follows. First, the parameters are estimated using an ARDL model, a very general form of which can be given as:

$$y_t, \Delta y_t = \alpha_0 + \delta T + \sum_{p=1}^P \phi_p y_{t-p} + \sum_{l_1=0}^{L_1} \theta_{1l_1} x_{1,t-l_1} + \cdots + \sum_{l_k=0}^{L_k} \theta_{kl} x_{t-l_k} + \sum_{m=1}^M \alpha_m \Delta y_{t-m} + \sum_{q_1=0}^{Q_1} \beta_{1q_1} \Delta x_{1,t-q_1} + \cdots + \sum_{q_k=0}^{Q_k} \beta_{kq_k} \Delta x_{k,t-q_k} + \epsilon_t \quad (1)$$

Here, the dependent variable—which can appear in either level form or first-differences—is a function of a constant, α_0 , a deterministic linear trend T , up to P lags of the dependent variable and L lags of the independent variables appearing in level form, up to M lags of the first-differenced dependent variable and up to Q lags of any first-differenced independent variables, and an independent and identically distributed error term ϵ_t . Note that L and Q may differ for each of the included k regressors. Users should use a combination of theory, information criteria, unit-root, cointegration (if applicable) and residual-based tests to arrive at a more restrictive specification of Equation 1 (De Boef & Keele, 2008; Philips, 2018; Wilkins, 2018).

Next, the estimated parameters are used to draw s simulations from a multivariate normal distribution, with mean $\hat{\beta}$ and variance obtained from the estimated variance-covariance matrix. Stable predicted values for y using the s simulations of $\hat{\beta}$ are then created, using starting values of the independent variables (usually means for continuous variables or modes for categorical ones). Users can choose from either expected values or—by incorporating fundamental model uncertainty—predicted values. Next, at a defined time period t , one of the independent variables is “shocked” by an amount determined by the researcher, and the effect on the dependent variable is observed across the resulting time periods. To summarize, models are estimated and counterfactual shocks are simulated through `dynardl()`. Then post-estimation dynamic inferences can be uncovered through the new command `dynardl.simulation.plot()`, which has six plotting responses mapped to the six visualizations that we outline below.

This functionality vastly simplifies the production and interpretation of dynamic models, increasing their applicability to a broader class of users. Dynamic models, like the ARDL model, can provide researchers with a variety of inferences about both short- and long-run effects of variables. Basic interpretation of dynamic models usually defaults to presenting both the short-run effect of a one-unit increase in an independent variable on the dependent variable, as well as the long-run, or cumulative total effect. The short-run effect is relatively easy to estimate and observe (represented by a $\hat{\beta}$ coefficient in an ARDL model), but long-run effects—and their associated confidence intervals—are harder to obtain, since they involve a non-linear combination of parameter estimates. As a result, they are often ignored, even though we are meant to be taking them “seriously” (De Boef & Keele, 2008).

In addition to the difficulties in calculating short- and long-run effects, inferences can also be obscured due to the complexity of the model. This is especially true as the models become more sophisticated (like including multiple lags and/or first-differences of dependent and independent variables). For short-run effects, increased model complexity encourages a focus on “boring” effects (the one-unit effect reported by the $\hat{\beta}$ in the regression output); for long-run effects, greater model complexity results in increasingly intractable closed-form estimates of the effects. Stochastic simulations can solve both of these problems by allowing for short-run dynamics beyond the traditional one-unit effect, and also by removing the need for analytic calculations when discussing long-run effects.

Plotting functionality

Specifically, `dynamac` implements six new visualizations. These can be presented one at a time using the command `dynardl.simulation.plot()`. Users wishing to present all six at once can do so with the command `dynardl.all.plots()`. The resulting plots are:

- The response path of the level of the dependent variable, given a change (a “shock”) in a single independent variable at a single point in time (Williams & Whitten, 2012). Note the value of this “shock” is user-defined, allowing for meaningful short-run inferences beyond the typical one-unit effect represented by a $\hat{\beta}$; for instance, a one standard deviation increase or decrease.
- Deviations between the predicted and average values of y_t (i.e., $\hat{y}_t - \bar{y}$). While this plot follows a response path identical to the plot above, if a shock dissipates and the series reverts to its mean, this is easier to observe if we do not have to subtract the stable starting value from the plot. If the series reverts to a value other than its mean, we would be especially interested in this new long-run mean in response to the shock. This is analogous to an impulse response function.
- The period-to-period response in y_t by differencing the response path. Such a figure allows us to visualize successive movements over time; if the change is equal to zero, it means that there is no movement between the past and current period.

- The size of the shock itself over time. This allows us to visualize dynamic persistence, so we can make statements like “how many time points until the shock is at 50 percent of its original magnitude?” or “how many time points until the shock is effectively zero?” This information is poorly represented by either the short-run or long-run effects as traditionally calculated, but is of interest to analysts.
- The cumulative response in y_t to a shock in x_t . This helps us to understand what the “final” effect of a shock to an independent variable does to a dependent variable. An independent variable might first cause a positive response, but ultimately may be overwhelmed by a negative movement as time goes on. Unlike the other plots, this considers the cumulative histories in each individual simulation when plotting the response in y over time, making the estimation of the confidence intervals more conservative. This is analogous to the traditional long-run effect of an ARDL model; however, it requires no analytical calculation.
- Total movement that occurs in y_t given a change in x_t . We might think of the “absolute” effect of a shock to x_t on y_t as being the sum over these positive and negative effects, but without allowing them to offset. In this setup, each of the period-over-period changes counts towards the “total” movement in y_t , so that all of the movement in y is attributed as an effect to the independent variable.

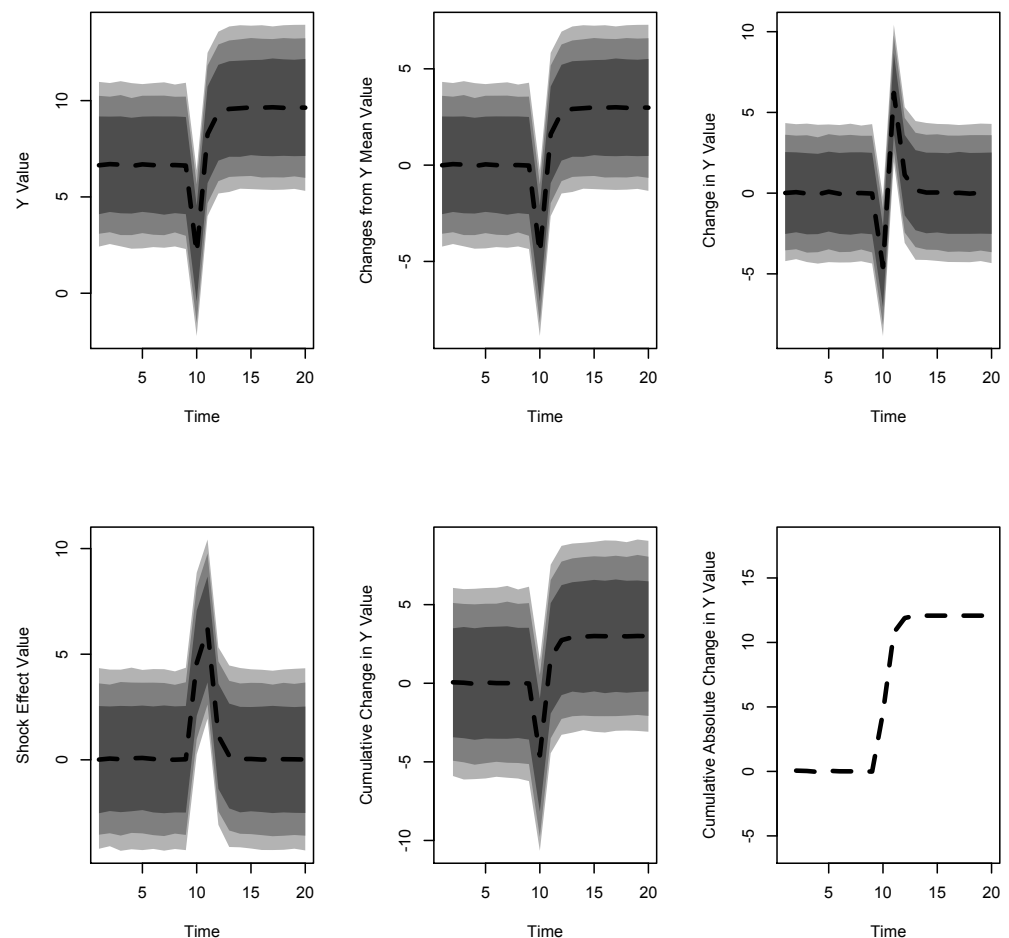


Figure 1: Six quantities of interest from the ARDL equation $\Delta y_t = -0.8y_{t-1} - 2\Delta x_t + x_{t-1} + u_t$.

Figure 1 plots these visualizations from a hypothetical ARDL model. Starting with the top-left plot and moving clockwise, users can easily understand how a dependent variable responds over time to a shock in an independent variable, how it changes away from its average pre-shock value, how the change itself changes over time, the cumulative absolute changes in the dependent variable, the cumulative nature of the changes in the dependent variable, and how the shock decays over time. All six graphical quantities of interest provide a richer set of information about our inferences.

The additional nicety is that this requires no additional mathematical burden on the user, regardless of the underlying complexity of the model. This is critical, as we are often encouraged to make modeling decisions that increase complexity, not realizing the resulting strain on analytic inferences. For instance, Wilkins (2018) advocates adding an additional lag of y_t at $t - 2$ in order to remove residual autocorrelation. Following this advice, the long-run effect is more difficult to calculate, as it is now a non-linear function of both coefficients on x_t at t and $t - 1$ as well as the coefficients on the lag of y_t and the additional lag at $t - 2$. Such complexity could reasonably encourage a practitioner to ignore long-run inferences; this is obviously not the preferred outcome. Through dynamac, though, stochastic simulations retrieve these quantities with no analytic calculations required. And the resulting graphic inferences are easy to both create and interpret, allowing users to focus on what they know best: their underlying models.

Conclusion

Dynamic models are complicated since they have many “hidden” effects that do not lend themselves to straightforward interpretation. Scholars that use these models typically interpret only the short- and long-run effects. Stochastic simulations can move our interpretations forward, even in the face of increased model complexity. As we have shown, there are six quantities of interest that are of interest to users, since they provide us with a richer understanding of the dynamic process taking place. dynamac now includes the ability to obtain all six of the visualizations described above in a single function: `dynardl.all.plots()`. In doing so, we help users expand the different types of quantities of interest they can use to assess statistical and substantive significance of their independent variables.

References

- Benton, A. L., & Philips, A. Q. (2020). Does the @realDonaldTrump really matter to financial markets? *American Journal of Political Science*, 64(1), 169–190. doi:[10.1111/ajps.12491](https://doi.org/10.1111/ajps.12491)
- Danish, R. U. (2020). Linking biomass energy and CO2 emissions in china using dynamic autoregressive-distributed lag simulations. *Journal of Cleaner Production*, 250, 119533. doi:[10.1016/j.jclepro.2019.119533](https://doi.org/10.1016/j.jclepro.2019.119533)
- De Boef, S., & Keele, L. (2008). Taking time seriously. *American Journal of Political Science*, 52(1), 184–200. doi:[10.1111/j.1540-5907.2007.00307.x](https://doi.org/10.1111/j.1540-5907.2007.00307.x)
- Demirhan, H. (2020). *dLagM: Time series regression models with distributed lag models*. Retrieved from <https://CRAN.R-project.org/package=dLagM>
- Jordan, S., & Philips, A. Q. (2018a). Cointegration testing and dynamic simulations of autoregressive distributed lag models. *The Stata Journal*, 18(4), 902–923. doi:[10.1177/1536867X1801800409](https://doi.org/10.1177/1536867X1801800409)
- Jordan, S., & Philips, A. Q. (2018b). Dynamic simulation and testing for single-equation cointegrating and stationary autoregressive distributed lag models. *The R Journal*, 10(2). doi:[10.32614/RJ-2018-076](https://doi.org/10.32614/RJ-2018-076)

- Jordan, S., & Philips, A. Q. (2020). *Dynamac: Dynamic simulation and testing for single-equation ARDL models*. Retrieved from <https://CRAN.R-project.org/package=dynamac>
- Khan, M. K., Teng, J.-Z., & Khan, M. I. (2019). Effect of energy consumption and economic growth on carbon dioxide emissions in Pakistan with dynamic ARDL simulations approach. *Environmental Science and Pollution Research*, 26, 23480–23490. doi:[10.1007/s11356-019-05640-x](https://doi.org/10.1007/s11356-019-05640-x)
- Liu, H., & Liang, S. (2019). The nexus between energy consumption, biodiversity, and economic growth in lancang-mekong cooperation (LMC): Evidence from cointegration and Granger causality tests. *International Journal of Environmental Research and Public Health*, 16(18), 3269. doi:[10.3390/ijerph16183269](https://doi.org/10.3390/ijerph16183269)
- Natsiopoulos, K., & Tzeremes, N. (2020). *ARDL: ARDL, ECM and bounds-test for cointegration*. Retrieved from <https://CRAN.R-project.org/package=ARDL>
- Philips, A. Q. (2018). Have your cake and eat it too? Cointegration and dynamic inference from autoregressive distributed lag models. *American Journal of Political Science*, 62(1), 230–244. doi:[10.1111/ajps.12318](https://doi.org/10.1111/ajps.12318)
- Sarkodie, S. A., Adams, S., Owusu, P. A., Leirvik, T., & Ozturk, I. (2020). Mitigating degradation and emissions in China: The role of environmental sustainability, human capital and renewable energy. *Science of the Total Environment*, 719, 137530. doi:[10.1016/j.scitotenv.2020.137530](https://doi.org/10.1016/j.scitotenv.2020.137530)
- Sarkodie, S. A., & Owusu, P. A. (2020). Investigating the cases of novel coronavirus disease (COVID-19) in china using dynamic statistical techniques. *Heliyon*, 6(4). doi:[10.1016/j.heliyon.2020.e03747](https://doi.org/10.1016/j.heliyon.2020.e03747)
- Sarkodie, S. A., Strezov, V., Weldekidan, H., Asamoah, E. F., Owusu, P. A., & Doyi, I. N. Y. (2019). Environmental sustainability assessment using dynamic autoregressive-distributed lag simulations—nexus between greenhouse gas emissions, biomass energy, food and economic growth. *Science of the Total Environment*, 668, 318–332. doi:[10.1016/j.scitotenv.2019.02.432](https://doi.org/10.1016/j.scitotenv.2019.02.432)
- Sharma, C. (2019). Testing the asymmetric effects of the economic policy uncertainty on the tourism demand in India. *Tourism Economics*. doi:[10.1177/1354816619894080](https://doi.org/10.1177/1354816619894080)
- Wilkins, A. S. (2018). To lag or not to lag?: Re-evaluating the use of lagged dependent variables in regression analysis. *Political Science Research and Methods*, 6(2), 393–411. doi:[10.1017/psrm.2017.4](https://doi.org/10.1017/psrm.2017.4)
- Williams, L. K., & Whitten, G. D. (2012). But wait, there's more! Maximizing substantive inferences from TSCS models. *The Journal of Politics*, 74(03), 685–693. doi:[10.1017/s0022381612000473](https://doi.org/10.1017/s0022381612000473)
- Zeileis, A. (2019). *Dynlm: Dynamic linear models and time series regression*. Retrieved from <https://CRAN.R-project.org/package=dynlm>