doi.org/10.26398/IJAS.0031-018

A FAMILY OF THEORY-BASED PARAMETRIC LAG SHAPES FOR DISTRIBUTED-LAG LINEAR REGRESSION

Alessandro Magrini[1]

Dep. Statistics, Computer Science, Applications – University of Florence, Italy

Abstract *Distributed-lag linear regression is a widely used tool in economic analysis when several explanatory variables are supposed to influence an outcome over several periods. In distributed-lag linear regression, the lag shape of an explanatory variable is the set of the coefficients associated to its temporally delayed (lagged) instances. Unfortunately, distributed-lag linear regression with no constraints on the lag shapes may entail problems of inference, due to potential multicollinearity and large number of free parameters, and of interpretation, since the lag shapes may have an irregular form and may not meet theoretical requirements. In this paper, we present type II constrained lag shapes, a new family of parametric lag shapes designed to represent theory-based lag structures through a small number of parameters. The proposed family improves the Almon's polynomial lag shape in terms of interpretability, and includes the well known endpoint-constrained quadratic and Gamma lag shapes. The applicability of the proposed family is illustrated through an economic impact assessment problem.*

Keywords: *distributed-lag linear regression; dynamic impact assessment; heteroskedasticity; stationary time series; temporal autocorrelation.*

1. INTRODUCTION

In economic analysis, a variable is often supposed to influence an outcome over several periods. Typical examples include the effect of public investments on gross value added, the effect of investments in research and development on the quality of commercial products, the effect of a change in income on consumption.

Linear regression with temporally delayed (lagged) instances of the explanatory variables, known as distributed-lag linear regression, is a widely used tool to address these kind of problems. In distributed-lag linear regression, the lag shape of an explanatory variable is the set of the coefficients associated to its lagged instances. The practical application of distributed-lag linear regression critically depends on the relationship among the coefficients composing the lag shape of each explanatory variable. From a technical point of view, multicollinearity easily occurs because the lagged instances of the same explanatory variable tend to be highly correlated, and the number of free parameters increases exponentially

[1]- Corresponding author: alessandro.magrini@unifi.it

as a new explanatory variable is considered, so that least squares estimators can have very high variance. From a theoretical point of view, the lag shape of an explanatory variable should have a regular form. For example, the effect of an investment may be small at first, then it may reach a peak before diminishing to zero after some time lags. In this view, lag shapes with no constraints can imply coefficients with different sign and/or following an irregular pattern, so that they may be difficult to interpret.

In this paper, we present *type II constrained lag shapes*, a new family of parametric lag shapes designed to represent theory-based lag structures through a small number of parameters. Type II constrained lag shapes improve the Almon's polynomial lag shape (Almon, 1965) in terms of interpretability, and include the well known endpoint-constrained quadratic (Andrews and Fair, 1992) and Gamma (Schmidt, 1974) lag shapes. They were recently employed to perform dynamic causal inference in Markovian structural causal models (Magrini, 2019b), and to assess the impact of Agricultural research expenditure (Magrini et al., 2019).

The paper is structured as follows. In Section 2, distributed-lag linear regression is presented. In Section 3, the family of type II constrained lag shapes is defined, several members are shown and their least squares estimation is addressed. In Section 4, the proposed methodology is applied to assess the impact of government expenditure and capital investments on international tourist arrivals in Northern Europe. Section 5 includes concluding remarks and considerations on possible future developments.

2. DISTRIBUTED-LAG LINEAR REGRESSION

Let Y be the outcome and $\boldsymbol{X} = (X_1, \dots, X_p)$ be a set of p explanatory variables. Also, let y_t and $x_{i,t}$ be the measurement of Y and X_i, respectively, at time t. We assume that: (i) time is discrete, (ii) the outcome and the explanatory variables are stationary, (iii) the influence of each explanatory variable on the outcome does not depend on time but only on the temporal distance (lag). Under these assumptions, distributed-lag linear regression is defined as:

$$y_t = \beta_0 + \sum_{i=1}^{p} \sum_{l=0}^{L_i} \beta_{i,l} x_{i,t-l} + \varepsilon_t \tag{1}$$

where $\beta_{i,l}$ is the coefficient for X_i at time lag l, L_i is the number of time lags of X_i being considered, and ε_t is the random error at time t.

The set $\boldsymbol{\beta}_i = (\beta_{i,0}, \beta_{i,1}, \dots, \beta_{i,L_i})$ is denoted as the *lag shape* of explanatory variable X_i and represents its influence on Y at different time lags. Note that the case where explanatory variable X_i influences the outcome Y only instantaneously is obtained by setting $L_i = 0$.

Lag shapes of the explanatory variables in model (1) can be consistently estimated using least squares provided that random errors are linearly uncorrelated with the explanatory variables. The violation of this condition, known as endogeneity, has several causes, like the omission of variables linearly correlated with both the outcome and some explanatory variables (confounders), the presence of measurement errors, the simultaneous determination of some variables (bi-directed causal relationships). Detecting and solving endogeneity goes beyond the scope of this paper, and it is not addressed here. A recent discussion of the topic can be found in Antonakis et al. (2014).

Besides consistency of least squares estimators, consistent estimation of their covariance matrix is also required to get reliable confidence intervals and significance tests for model (1). Consistent estimation of the covariance matrix of least squares estimatiors critically depends on the assumptions on the random errors. Standard assumptions like independence and homoskedasticity hardly apply to time series, because both the correlation between the random errors and their variance may depend on time. A practical solution is to fit the regression model (1) using least squares estimation, and then to compute the Heteroskedasticity and Autocorrelation Consistent (HAC) covariance matrix of least squares estimators (Newey and West, 1978).

3. TYPE II CONSTRAINED LAG SHAPES

As discussed above, the regression model (1) may entail several practical problems, like multicollinearity, high number of free parameters and irregular lag shapes. The Almon's polynomial lag shape (Almon, 1965) addresses the problem of multicollinearity and large number of free parameters by forcing the coefficients to follow a polynomial of order Q:

$$\beta_{i,l} = \sum_{q=0}^{Q} \phi_{i,q}\, l^q \tag{2}$$

For instance, for $Q = 2$ we have $\beta_{i,l} = \phi_{i,0} + \phi_{i,1} l + \phi_{i,2} l^2$. However, the Almon's polynomial lag shape may still imply multiple modes and coefficients with different signs, which can be difficult to interpret. The endpoint-constrained quadratic (Andrews and Fair, 1992) and the Gamma (Schmidt, 1974) lag shapes improve the Almon's lag shape in terms of interpretability: the former forces coefficients to follow a second order polynomial with endpoints constrained to value 0, while the latter imposes an asymmetric non-polynomial function. Starting from these two lag shapes, we designed *type II constrained lag shapes*, a new family of parametric lag shapes suitable to represent theory-based lag structures through a small number of parameters. In the following subsections, type II constrained lag shapes are defined, some notable members are illustrated and their least squares estimation is addressed.

3.1 DEFINITION

A *type II constrained lag shape* for explanatory variable X_i is defined as:

$$\beta_{i,l} = \theta_i \, w(l; a_i, b_i) \qquad l = 0, \ldots, L_i \tag{3}$$

where θ_i is the *scale* parameter and $w(l; a_i, b_i)$ is a function of the time lag mapping to the interval $[0,1]$, called *weight function*, which depends on two parameters a_i and b_i, called *shape parameters*. The scale parameter determines the maximum effect of X_i on Y at any time lag, while the weight function establishes how the effect is distributed across the time lags. A type II constrained lag shape is a reduced-dimensional representation of a lag shape: whichever the maximum time lag L_i considered, the lag shape is completely defined by one scale parameter and by two shape parameters, as indicated by the term 'type II'. An important property is that the scale parameter defines the sign of the whole lag shape:

$$\begin{array}{l} \beta_{i,l} > 0 \Longleftrightarrow \theta_i > 0 \\ \beta_{i,l} < 0 \Longleftrightarrow \theta_i < 0 \end{array} \qquad \forall l : w(l; a_i, b_i) \neq 0 \tag{4}$$

called *sign monotonicity*.

Consider, as an example, a lag shape decreasing linearly from a maximum effect θ_i at lag a_i to zero at lag b_i. This lag shape, to which we refer as *linearly decreasing* (LD) lag shape, has weight function:

$$w_{\mathrm{LD}}(l; a_i, b_i) = \frac{b_i - l}{b_i - a_i} \mathscr{I}_{a_i \leq l \leq b_i} \tag{5}$$

where $\mathscr{I}$ is the indicator function.

3.2 SOME TYPE II CONSTRAINED LAG SHAPES

As stated above, the *endpoint-constrained quadratic* (ECQ) lag shape (Andrews and Fair, 1992) belongs to the family of type II constrained lag shapes. It has weight function:

$$w_{\text{ECQ}}(l;a_i,b_i) = \left[-\frac{4}{(b_i-a_i+2)^2}l^2 + \frac{4(a_i+b_i)}{(b_i-a_i+2)^2}l - \frac{4(a_i-1)(b_i+1)}{(b_i-a_i+2)^2}\right]\mathscr{I}_{a_i\le l\le b_i}$$
$$a_i,b_i \in \mathbb{N} \qquad a_i \le b_i \tag{6}$$

Basically, the ECQ lag shape is an Almon's second-order polynomial lag shape, but it is constrained to zero for $l \le a_i - 1$ or $l \ge b_i + 1$. As such, it is symmetric and has mode equal to θ_i at lag $(a_i + b_i)/2$.

The *quadratic decreasing* (QD) lag shape is the second-order polynomial version of the linearly decreasing (LD) lag shape shown in Subsection 3.1:

$$w_{\text{QD}}(l;a_i,b_i) = \left[\frac{l^2-2(b_i+1)l+(b_i+1)^2}{(b_i-a_i+1)^2}\right]\mathscr{I}_{a_i\le l\le b_i} \qquad a_i,b_i \in \mathbb{N} \qquad a_i \le b_i \tag{7}$$

The QD lag shape decreases from value θ_i at lag a_i to value 0 at lag $b_i + 1$.

The *Gamma* lag shape (Schmidt, 1974) is another well known lag shape belonging to the family of type II constrained lag shapes. It has weight function:

$$w_{\text{GAM}}(l;a_i,b_i) = (l+1)^{\frac{a_i}{1-a_i}} b_i^l \left[\left(\frac{a_i}{(a_i-1)\log b_i}\right)^{\frac{a_i}{1-a_i}} b_i^{\frac{a_i}{(a_i-1)\log b_i}}\right]^{-1} \qquad 0 < a_i < 1 \qquad 0 < b_i < 1 \tag{8}$$

The Gamma lag shape is different from zero at any time lag, and is positively skewed with mode equal to θ_i at lag $\frac{a_i}{(a_i-1)\log b_i}$.

Note that, the shape parameters of the ECQ, QD and LD lag shapes represent the endpoint of an interval of time lags where the effect of explanatory variable X_i is not zero. Specifically, a_i is the *gestation lag*, b_i is the *lead lag*, and $b_i - a_i$ is the *lag width*. As such, the ECQ, QD and LD lag shapes degenerate into an instantaneous effect if $a_i = b_i = 0$. Instead, the shape parameters of the Gamma lag shape are not the gestation and the lead lags, and, since the Gamma lag shape is different from zero at any time lag, it cannot reduce exactly to an instantaneous effect. Note that, for the Gamma lag shape, the gestation lag is equal to zero and the lead lag is not properly defined, but it may be approximated by the first time lag after the mode where the weight function is lower than a tolerance threshold, for example 10^{-4}.

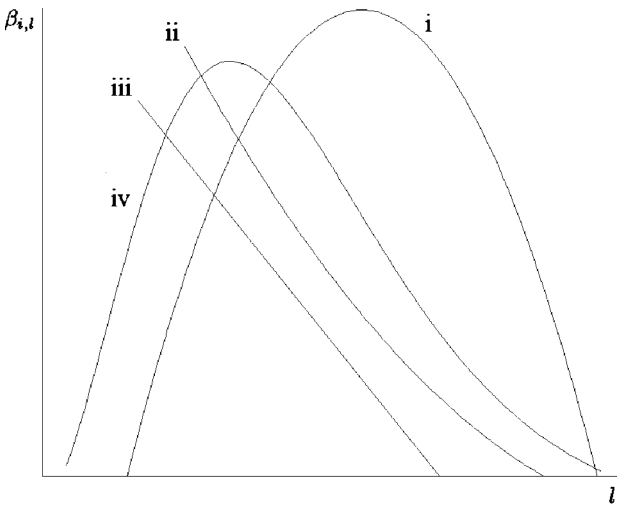

Fig. 1: Examples of type II constrained lag shapes: i) endpoint-constrained quadratic (ECQ); ii) quadratic decreasing (QD); iii) linearly decreasing (LD); iv) Gamma.

3.3 LEAST SQUARES ESTIMATION OF TYPE II CONSTRAINED LAG SHAPES

By replacing each $\beta_{i,l}$ in the regression model (1) with a type II constrained lag shape, we get:

$$y_t = \beta_0 + \sum_{i=1}^{p} \sum_{l=0}^{L_i} \theta_i \, w(l; a_i, b_i) \, x_{i,t-l} + \varepsilon_t \tag{9}$$

which can be written as:

$$y_t = \beta_0 + \sum_{i=1}^{p} \theta_i x^*_{i,t} + \varepsilon_t \tag{10}$$

where:

$$x^*_{i,t} = \sum_{l=0}^{L_i} w(l; a_i, b_i) \, x_{i,t-l}$$

Thus, if the shape parameters a_i and b_i are known for $i = 1, \ldots, p$, model (1) becomes a classic linear regression where the explanatory variables are the transformed variables $X^*_1, \ldots, X^*_p$. As such, least squares estimation can be applied to estimate the scale parameters $\theta_1, \ldots, \theta_p$. Once the regression model is fitted, the HAC covariance matrix is computed and the standard error (SE) of the estimates of each scale parameter is derived. Note that the standard error of the estimate of

$\beta_{i,l}$, i.e., the coefficient for X_i at lag l, is:

$$\mathrm{SE}(\widehat{\beta}_{i,l}) = |w(l; a_i, b_i)| \ \mathrm{SE}(\theta_i) \tag{11}$$

In real-world applications, the shape parameters are unknown, thus we have to infer them from data. In general, the lag shape of variable X_i can be straightforwardly inferred from data if the shape parameters of all other explanatory variables are known. For instance, it is sufficient to fit several regression models by varying the pairs of the shape parameters for X_i, and to select the one with the minimum mean squared error (MSE):

$$\mathrm{MSE}_{t*} = \frac{1}{T - t^*} \sum_{t=t^*+1}^{T} (y_t - \widehat{y}_t)^2 \tag{12}$$

where $\widehat{y}_t$ is the value of Y at time t predicted by the regression model, T is the total number of periods and t^* is the maximum among the lead lags implied by all the pairs of shape parameters for X_i to be tested. Only the periods after t^* are considered in the computation in order to make the MSE comparable across all the competing regression models.

The pairs of shape parameters for X_i to be tested can be either all the possible ones if no prior knowledge is available, or a subset of the possible ones compatible with prior knowledge. For instance, it is often the case that the researcher knows that the gestation lag is comprised between two values, or that the lead lag is at most a certain value, or that the lag width is at least a certain value. Thus, the pairs of shape parameters not satisfying prior knowledge can be disregarded. The procedure of inferring the shape parameters from data is denoted as *adaptation,* and it is shown as Algorithm 1.

Note that the researcher may know the type II constrained lag shape for an explanatory variable (e.g., an ECQ lag shape) and be uncertain on the shape parameters, but also may be uncertain on the type II constrained lag shape itself (e.g., a Gamma, rather than a ECQ lag shape). The matrix PAR.TEST in Algorithm 1 will include, in the first case, pairs of shape parameters of a single type II constrained lag shape, while, in the second case, pairs of shape parameters of several type II constrained lag shapes.

The procedure shown as Algorithm 1 assumes that the shape parameters of all other explanatory variables are known. Unless $p = 1$, this is seldom the case in real-world problems, and, if one wants to adapt $p > 1$ lag shapes altogether, the number of regression models to be fitted would increase exponentially with p. Algorithm 2 is a heuristic solution to jointly adapt $p > 1$ lag shapes with linear complexity in p.

Algorithm 1 Adaptation of one lag shape

Let X_i be the explanatory variable its lag shape should be adapted.

Require. A matrix PAR.TEST containing in each row a pair of shape parameters for X_i to be tested. For $j = 1,\dots,p\backslash i$, a vector $\mathsf{PAR}^{(j)}$ including the (known) shape parameters of X_j.

1. Initialize MSE equal to $+\infty$. For $j = 1,\dots,p\backslash i$, set $\mathsf{PAR.OK}^{(j)}$ equal to $\mathsf{PAR}^{(j)}$.
2. For each row in PAR.TEST:
   - set $\mathsf{PAR}^{(i)}$ equal to the values in the current row of PAR.TEST;
   - fit the distributed-lag linear regression with shape parameters as in vectors $\mathsf{PAR}^{(1)},\dots,\mathsf{PAR}^{(p)}$;
   - if the mean squared error is less than MSE, then:
     - set MSE equal to the mean squared error;
     - set $\mathsf{PAR.OK}^{(i)}$ equal to $\mathsf{PAR}^{(i)}$;
3. Return MSE and $\mathsf{PAR.OK}^{(i)}$.

Algorithm 2 Adaptation of $p > 1$ lag shapes

Require. For $i = 1,\dots,p$, a matrix $\mathsf{PAR.TEST}^{(i)}$ containing in each row a pair of shape parameters for X_i to be tested.

1. For $i = 1,\dots,p$, set $\mathsf{PAR}^{(i)}$ equal to a pair of shape parameters implying a gestation lag equal to the minimum among the gestation lags for X_i to be tested, and lead lag equal to the maximum among the lead lags for X_i to be tested. Set $\mathsf{X.TEST} = \{1,\dots,p\}$.
2. Repeat until X.TEST is not empty:
   - initialize MSE.TEMP as an empty vector of length p. For $i = 1,\dots,p$, set $\mathsf{PAR.TEMP}^{(i)}$ equal to $\mathsf{PAR}^{(i)}$;
   - for each value i in X.TEST:
     - Adapt the lag shape of X_i using Algorithm 1 with inputs the pairs of shape parameters in $\mathsf{PAR.TEST}^{(i)}$ and the lag shapes as in vectors $\mathsf{PAR.TEMP}^{(1)},\dots,\mathsf{PAR.TEMP}^{(p)}$. Set the i-th position of MSE.TEMP equal to the resulting mean squared error, and $\mathsf{PAR.TEMP}^{(i)}$ equal to the resulting shape parameters;
   - set $\mathsf{PAR}^{(h)}$ equal to $\mathsf{PAR.TEMP}^{(h)}$ and remove value h from X.TEST, where h is the position with minimum value in MSE.TEMP.
3. Return $\mathsf{PAR}^{(1)},\dots,\mathsf{PAR}^{(p)}$.

4. READ–WORLD APPLICATION

The proposed family of lag shapes is applied to assess the impact of government expenditure and capital investments on international tourist arrivals in Northern Europe. We considered the following countries: Denmark, Estonia, Iceland, Finland, Latvia, Norway and Sweden in the period 1995–2016, downloading yearly data from the World Tourism & Travel Council and the Eurostat databases.

We included several potential confounders: population, gross domestic product, purchase power parity, employment in tertiary sector to total employment, number of world heritage sites and a dummy variable for the economic crisis until 2012. We did not include data on tourism infrastructures, as they are an intermediate product of the causal chain linking government expenditure and capital investments to international arrivals. All variables were taken in log-return, excepting the number of world heritage sites, which was taken in first order difference, and the dummy for the economic crisis, which was not transformed. Stationarity of all the time series was confirmed by the Kwiatkowski-Phillips-Schmidt-Shin (KPSS, Kwiatkowski et al., 1992) test with *p*-values of the different countries combined according to the method by Demetrescu et al. (2006). Descriptive statistics are shown in Table 1, while Figure 2 shows the median log-return by year of international arrivals, government expenditure and capital investments.

Fixed intercepts for the countries were assumed in order to take into account the panel structure of data. Algorithm 2 was applied to adapt the lag shapes of government expenditure and capital investments on international arrivals. In the adaptation, we considered ECQ, QD, LD and Gamma lag shapes with maximum gestation lag of 5 years, minimum lag width of 3 years and maximum lead lag of 15 years. Standard errors were computed according to HAC estimation of the covariance matrix of least squares estimators, with serial correlation order inferred for each country based on the Akaike information criterion (AIC). The results are shown in Tables 2-3 and Figure 3. All the computations and the graphics were obtained using the `R` package `dlsem` (Magrini, 2019a).

Results show that the lag shape with the best fit to data for government expenditure is $\text{ECQ}(0.2602, 0, 8)$, while for capital investments is $\text{ECQ}(0.2492, 1, 14)$. Both these two lag shapes have a statistically significant scale parameter. Note that a statistically significant scale parameter implies that all the coefficients in the lag shape are statistically significant, as it is apparent from confidence intervals in Figure 3. The estimated lag shapes suggest that the effect of government expenditure on international arrivals has zero year of gestation and lasts up to 8 years, while the effect of capital investments has one year of gestation and lasts up to 14 years.

The statistically significant and positive coefficients for population and power purchasing parity indicate that both the propensity to spend or invest and the volume of international tourists depend directly on the size and on the price level of a country. Instead, the coefficient of gross domestic product did not result statistically significant, suggesting that the economic development level is not a confounder for the investigated relationships once population and power purchasing parity are taken into account. Also, the coefficient for the dummy for economic crisis resulted statistically significant and negative, pointing out that the level of tourism volume in the considered countries was lower before 2012. Finally, the statistically significant intercepts for Norway and Sweden indicate a lower tourism volume for these two countries at constant government expenditure, capital investments, population, power purchasing parity and general economic situation (crisis or not).

Tab. 1: Descriptive statistics for the tourism data in Northern Europe countries, 1995–2016.

Variables in level	Minimum	1st quartile	Median	3rd quartile	Maximum
International arrivals ($\times 10^3$)	190.0	1194.0	2313.5	4270.0	10781.0
Government expenditure ($\times 10^6$ USD)	10.0	30.0	132.2	237.9	677.7
Capital investments ($\times 10^6$ USD)	30.0	316.1	1231.8	2326.8	4472.1
Population ($\times 10^3$ people)	267.5	138.9	4642.0	5411.7	9903.1
Gross domestic product ($\times 10^9$ USD)	4.4	16.9	161.8	280.6	578.7
Purchasing power parity	0.3	0.5	8.4	9.2	141.0
Employment in tertiary sector (%)	52.4	65.8	71.2	75.7	79.7
Number of world heritage sites	0.0	1.0	4.0	6.0	12.0
Variables in return	Minimum	1st quartile	Median	3rd quartile	Maximum
International arrivals	0.786	1.003	1.045	1.102	1.475
Government expenditure	0.337	1.000	1.000	1.034	2.707
Capital investments	0.298	0.921	1.048	1.182	4.770
Population	0.979	0.998	1.004	1.007	1.026
Gross domestic product	0.731	0.994	1.052	1.129	1.441
Purchasing power parity	0.904	0.991	1.005	1.019	1.226
Employment in tertiary sector (%)	0.944	1.001	1.004	1.013	1.060
Variables in log-return	Minimum	1st quartile	Median	3rd quartile	Maximum
International arrivals	−0.241	0.003	0.044	0.097	0.389
Government expenditure	−1.088	0.000	0.000	0.034	0.996
Capital investments	−1.211	−0.082	0.047	0.168	1.562
Population	−0.021	−0.002	0.004	0.007	0.025
Gross domestic product	−0.314	−0.006	0.051	0.122	0.365
Purchasing power parity	−0.100	−0.009	0.005	0.019	0.204
Employment in tertiary sector (%)	−0.057	0.001	0.004	0.012	0.058

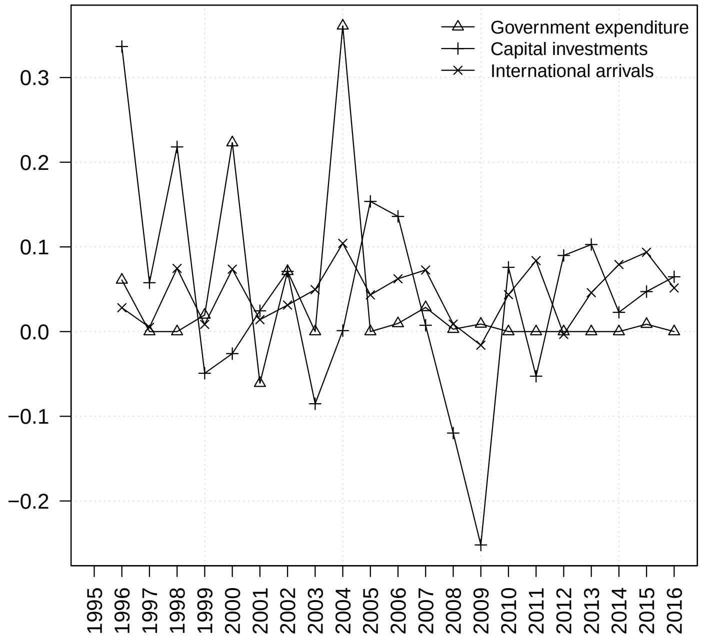

Fig. 2: International arrivals, government expenditure and capital investments in Northen Europe countries, 1995-2016 (median log-return by year).

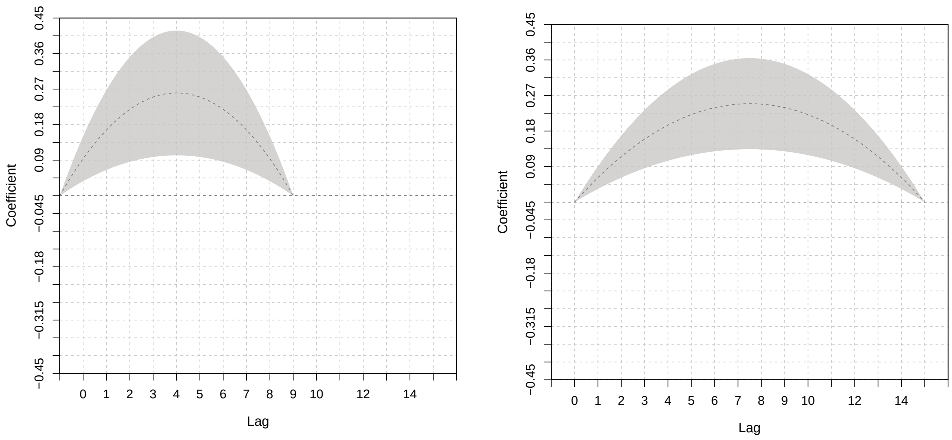

Fig. 3: Estimated lag shape of government expenditure (left panel) and capital investments (right panel) on international tourist arrivals in Northern Europe. Coefficients represent percentage variations. Shaded regions represent asymptotic 95% confidence intervals, where standard errors are based on HAC estimation of the covariance matrix of least squares estimators.

Tab. 2: Parameter estimation of the distributed-lag linear regression of international tourist arrivals from government expenditure and capital investments in Northern Europe. Lag shapes were adapted using Algorithm 2. Standard errors are based on HAC estimation of the covariance matrix of least squares estimators. *p*-values are based on a t-statistic with 34 degrees of freedom. Multiple R-squared is 0.755.

Outcome variable			
International tourist arrivals (log-return)			
Explanatory variables	**Estimate**	**SE**	***p*-value**
Government expenditure (log-return)	ECQ(0.260, 0, 8)	0.081	0.003
Capital investments (log-return)	ECQ(0.249, 1, 14)	0.059	0.000
Potential confounders	**Estimate**	**SE**	***p*-value**
Population (log-return)	14.422	3.615	0.000
Gross domestic product (log-return)	0.173	0.122	0.166
Purchasing power parity (log-return)	0.926	0.401	0.027
Employment in tertiary sector (log-return)	–0.854	0.753	0.265
Number of world heritage sites (1st order difference)	0.039	0.033	0.240
Dummy for economic crisis	–0.054	0.026	0.045
Intercepts	**Estimate**	**SE**	***p*-value**
Denmark	–0.051	0.039	0.192
Estonia	–0.044	0.032	0.177
Finland	–0.059	0.029	0.052
Iceland	–0.075	0.055	0.182
Latvia	0.024	0.084	0.780
Norway	–0.283	0.067	0.000
Sweden	–0.221	0.055	0.000

Tab. 3: Estimated lag shapes of government expenditure and capital investments on international tourist arrivals in Northern Europe. Coefficients represent percentage variations. The cumulative coefficient of an explanatory variable at lag *l* represent its effect up to *l* time lags. Standard errors are shown within brackets.

	Government expenditure		Capital investments	
Lag	Coefficient	Cumulative coefficient	Coefficient	Cumulative coefficient
0	0.094 (0.029)	0.094 (0.029)	0.000 (0.000)	0.000 (0.000)
1	0.167 (0.052)	0.260 (0.059)	0.062 (0.015)	0.062 (0.015)
2	0.219 (0.068)	0.479 (0.090)	0.115 (0.027)	0.177 (0.031)
3	0.250 (0.077)	0.729 (0.119)	0.160 (0.038)	0.337 (0.049)
4	0.260 (0.081)	0.989 (0.144)	0.195 (0.046)	0.532 (0.067)
5	0.250 (0.077)	1.239 (0.163)	0.222 (0.052)	0.753 (0.085)
6	0.219 (0.068)	1.457 (0.177)	0.239 (0.057)	0.993 (0.102)
7	0.167 (0.052)	1.624 (0.184)	0.248 (0.059)	1.241 (0.118)
8	0.094 (0.029)	1.718 (0.186)	0.248 (0.059)	1.489 (0.131)
9	0.000 (0.000)	1.718 (0.186)	0.239 (0.057)	1.728 (0.143)
10	0.000 (0.000)	1.718 (0.186)	0.222 (0.052)	1.950 (0.152)
11	0.000 (0.000)	1.718 (0.186)	0.195 (0.046)	2.145 (0.159)
12	0.000 (0.000)	1.718 (0.186)	0.160 (0.038)	2.304 (0.164)
13	0.000 (0.000)	1.718 (0.186)	0.115 (0.027)	2.419 (0.166)
14	0.000 (0.000)	1.718 (0.186)	0.062 (0.015)	2.481 (0.166)
15	0.000 (0.000)	1.718 (0.186)	0.000 (0.000)	2.481 (0.166)

The cumulative coefficients at lag 14 were estimated as 1.718 and 2.481. Since the variables are in log-return, this means that, after 14 years, international arrivals are expected to increase approximatively by 1.7% for a unit percentage increase in government expenditure at constant capital investments, and by 2.5% for a unit percentage increase in capital investments at constant government expenditure.

5. CONCLUDING REMARKS

Distributed-lag linear regression is challenging in economic analysis, because the explanatory variables may have a long lead lag and the lag shapes should fit to some theoretical requirements. Unconstrained lag shapes may entail both problems of inference, due to potential multicollinearity and large number of free parameters, and of interpretation, due to the occurrence of multiple modes and coefficients with different signs. The Almon's polynomial lag shape is a well-known parametric lag shape which may solve the problems of inference, but not of interpretation. The family of type II constrained lag shapes presented in this paper is an improvement of the Almon's polynomial lag shape, also suited to represent theory-based lag structures. A type II constrained lag shape is defined by a scale parameter, determining the maximum effect of the explanatory variable at any time lag, and by two shape parameters, establishing how the effect is distributed across the time lags. This way, the problems of multicollinearity and large number of free parameters are completely overcome. Also, a type II constrained lag shape is characterized by parameters with clear economic interpretation. For instance, the gestation lag, the lead lag and the lag width can be explicit parameters.

The family of type II constrained lag shapes includes the well known endpoint-constrained quadratic and Gamma lag shapes. In this paper, we illustrated two further members of the family, but several other members can be defined at convenience.

Least squares estimation of type II constrained lag shapes is straightforward if all the shape parameters are known, or if the researcher is uncertain on a small set of them. Instead, if the shape parameters are unknown, it is required to fit a number of regression models increasing exponentially with the number p of explanatory variables. A heuristic procedure with linear complexity in p was proposed, and future work could be directed towards its improvement.

In order to achieve further flexibility, the family of type II constrained lag shapes can be extended by allowing more than two shape parameters. In this case, the proposed estimation procedure still applies, but k-uples of shape parameters, with with $k > 2$, should be considered.

Numerical maximization of the log-likelihood is an alternative to least squares estimation worthy of exploring in the future, as all the shape and the scale parameters could be estimated at one time.

REFERENCES

Almon, S. (1965). The Distributed Lag between Capital Appropriations and Net Expenditures. In *Econometrica*, 33: 178–196.

Andrews, W.K. and Fair, R.C. (1992). Estimation of Polynomial Distributed Lags and Leads with End Point Constraints. In *Journal of Econometrics*, 53: 123– 139.

Antonakis, J., Bendahan, S., Jacquart, P. and Lalive, R. (2014). Causality and Endogeneity: Problems and Solutions. In D.V. Day, ed., *The Oxford Handbook of Leadership and Organizations*. Oxford University Press, Oxford, UK.

Demetrescu, M., Hassler, U. and Tarcolea, A. (2006). Combining Significance of Correlated Statistics with Application to Panel Data. In *Oxford Bulletin of Economics and Statistics*, 68 (5): 647–663.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. In *Journal of Econometrics*, 54 (1-3): 159–178.

Magrini, A. (2019a). Distributed-Lag Linear Structural Equation Models in R: the dlsem Package. In *Austrian Journal of Statistics*, 48: 14–42.

Magrini, A. (2019b). Linear Markovian Models for Lag Exposure Assessment. In *Biometrical Letters*, 55 (2): 179–195.

Magrini, A., Bartolini, F., Coli, A. and Pacini, B. (2019). A Structural Equation Model to Assess the Impact of Agricultural Research Expenditure on Multiple Dimensions. In *Quality & Quantity*, 53 (4): 2063–2080.

Newey, W.K. and West, K.D. (1978). A Simple, Positive Semi-Definite, Het- eroskedasticity and Autocorrelation Consistent Covariance Matrix. In *Econo- metrica*, 55 (3): 703–708.

Schmidt, P. (1974). A Modification of the Almon Distributed Lag. In *Journal of the American Statistical Association*, 69: 679–681.