On the econometrics of the Koyck model*

Philip Hans Franses and Rutger van Oest

Econometric Institute, Erasmus University Rotterdam

P.O. Box 1738, NL-3000 DR, Rotterdam, The Netherlands

Econometric Institute Report 2004-07

Abstract

The geometric distributed lag model, after application of the so-called Koyck transformation, is often used to establish the dynamic link between sales and advertising. This year, the Koyck model celebrates its 50th anniversary.

In this paper we focus on the econometrics of this popular model, and we show that this seemingly simple model is a little more complicated than we always tend to think. First, the Koyck transformation entails a parameter restriction, which should not be overlooked for efficiency reasons. Second, the t-statistic for the parameter for direct advertising effects has a non-standard distribution. We provide solutions to these two issues.

For the monthly Lydia Pinkham data, it is shown that various practical decisions lead to very different conclusions.

Key words and phrases: Koyck model, sales-advertising relationship

---

*We thank Dennis Fok, Richard Paap and Marco Vriens for helpful discussions, and Lotje Kruithof for excellent research assistance. The address for correspondence is Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, email: franses@few.eur.nl or vanoest@few.eur.nl. The computer programs used in this paper are available from the second author.

1 Introduction

The geometric distributed lag model is often used to investigate the current and carryover effect of advertising on sales. This model makes current sales a function of current and past advertising levels, where the lag coefficients have a geometrically decaying pattern. As this model involves an infinite number of lagged variables, one often considers the so-called Koyck transformation (Koyck, 1954). In many studies the resultant model is hence called the Koyck model. Leendert Marinus Koyck (1918-1962) was a Dutch economist who studied and worked at the Netherlands School of Economics, which is now called the Erasmus University Rotterdam.

In this research note we will discuss the basic Koyck model, and illustrate that this model is less straightforward to analyze than is usually assumed or suggested. We will provide a discussion of possible solutions. Next, we will show for the well-known monthly Lydia Pinkham data that various approaches lead to different conclusions, thereby emphasizing the relevance of the proper methods.

2 The Koyck model

Consider the variables sales S_t and advertising A_t, and assume that the link between these two variables is given by

$$S_t = \mu + \beta(A_t + \lambda A_{t-1} + \lambda^2 A_{t-2} + ...) + \varepsilon_t, \tag{1}$$

where ε_t is an uncorrelated error variable. The parameter λ is usually called the retention rate. The current effect of advertising is β, whereas the carryover effect is equal to $\frac{\beta}{1-\lambda}$. To provide some intuition for the discussion below, note that when $\beta = 0$, the parameter λ disappears from the model, and cannot be retrieved. From a marketing perspective this is no problem as the model assumes that the current and carryover effects of advertising are both a function of β. However, from a statistical perspective, matters become a little complicated.

As this model contains a very large number of variables, it is common practice to apply the Koyck transformation. This entails subtracting λS_{t-1} from (1), to get

$$S_t = \mu + \beta A_t + \lambda S_{t-1} + \varepsilon_t - \lambda\varepsilon_{t-1}. \tag{2}$$

In time series jargon, this model is called an ARMAX model, see Franses (1991) for more details on ARMAX models. The autoregressive [AR] part concerns S_{t-1}, the moving average part [MA] concerns ε_{t-1} and the explanatory variables part [X] concerns A_t. Note that the parameter λ appears twice, and hence that, except for the intercept, there are only two parameters to estimate, while the model effectively contains three explanatory variables. Additionally, note that when $\beta = 0$, the model contains the lag polynomial $1 - \lambda L$, with L the lag operator, on both sides, which gets cancelled, that is, when $\beta = 0$, the model reduces to $S_t = \mu + \varepsilon_t$.

Parameter estimation

There are several approaches that one can follow to estimate the parameters in the resulting Koyck model in (2). Of course, the appropriate estimation method here is the maximum likelihood method, which imposes that the AR and MA parameter are the same. The (conditional) log-likelihood function is given by

$$\ln L(\mu, \beta, \lambda, \sigma^2) = -\frac{T-1}{2}\Big(\ln(2\pi) + \ln(\sigma^2)\Big) - \sum_{t=2}^{T} \frac{\varepsilon_t^2}{2\sigma^2}\,, \tag{3}$$

where T denotes the number of observations, and the $\{\varepsilon_t\}$ are recursively defined as

$$\begin{aligned} \varepsilon_1 &= 0, \\ \varepsilon_t &= S_t - \mu - \beta A_t - \lambda S_{t-1} + \lambda \varepsilon_{t-1}, \qquad t = 2, \ldots, T. \end{aligned} \tag{4}$$

This approach is also described in Hamilton (1994, p. 132) for general ARMA models. Asymptotic standard errors are obtained by taking the square roots of the diagonal elements of the estimated covariance matrix, which in turn can be computed as minus the inverse of the Hessian of (3) evaluated for the optimal parameter values. Numerical techniques, such as the BFGS algorithm or the Newton-Raphson algorithm, have to be used to get the maximum likelihood parameter estimates.

It is tempting though to decide to all the way neglect the MA part, so that the model parameters can be estimated using the method of ordinary least squares. This is obviously not a sound approach, as ε_{t-1} and S_{t-1} are not uncorrelated, and thereby one of the basic premises of regression theory gets violated. Simulation results in Table 1 suggest that the resultant downward bias of $\hat{\lambda}$ can be substantial.

One can also decide to estimate the parameters in an unrestricted ARMAX model, that is,

$$S_t = \mu + \beta A_t + \lambda_1 S_{t-1} + \varepsilon_t - \lambda_2 \varepsilon_{t-1}. \tag{5}$$

It is possible to get estimators for λ_1 and for λ_2 which are consistent estimators for λ. In practice, however, it is most likely that the corresponding estimates take different values, and the question appears which one should take. Also, (5) cannot be transformed back to a model like (1), and hence is less interesting from a theoretical perspective. Moreover, correctly imposing that $\lambda_1 = \lambda_2$ yields a more efficient estimator. The simulation results in Table 1 suggest that estimating λ_1 in an unrestricted ARMAX model leads to almost no bias.

Testing advertising effects

A second issue of concern for the Koyck model is a test for the significance of the advertising effects. Indeed, one may want to examine whether β is equal to zero. This is not trivial as under the null hypothesis of interest, that is, $\beta = 0$, the parameter λ disappears from the model, see (1) and (2). This is what is known as the Davies (1977) problem, and it seriously complicates statistical analysis. The issue is that the usually considered t-statistic depends on λ for which it is not clear which value to take. An appealing approach might seem to simply set λ at its maximum likelihood value. However, this would make the test (and its critical values) dependent on the data, and the asymptotic distribution would not be standard normal.

Recent solutions to the Davies problem are provided by Andrews and Ploberger (1994) and Hansen (1996), see also Carrasco (2002). The main idea is that one constructs a new test statistic based on the entire distribution of the original test statistic over a range of values of the unidentified parameter λ. In the Koyck model, a sensible range for λ would be the interval $[0, 1)$. One possibility, involving the entire distribution over λ, would be to consider the class of "sup test statistics", which corresponds to the highest value of the original test statistic within the range for λ. This approach is advocated by Davies (1977), see also Hansen (1996) and Carrasco (2002). Alternatively, one can consider the class of "ave test statistics", based on the average value of the original test statistic. This approach is put forward by Andrews and Ploberger (1994), and is further investigated by Hansen (1996). In each case, the asymptotic distribution of the resulting test statistic is not standard normal, so that its distribution has to be simulated.

In this paper we consider the "ave" and "sup" versions of the absolute t statistic $|t_\beta|$ and the Wald statistic t_β^2, where t_β is obtained by taking the ratio of the maximum likelihood estimate of β and its estimated asymptotic standard error. So, we focus on four test statistics, that is, ave absolute t, ave Wald, sup absolute t, and sup Wald. Although the two sup tests are equivalent, we include them both in order to achieve symmetry. Table 2 contains the simulated (asymptotic) critical values for the four tests at confidence levels of 80%, 90%, 95%, and 99%. In order to obtain these critical values, we ran 40000 simulations for $T = 1000$ observations. In each simulation, advertising data were drawn from a standard normal distribution. Next, sales data were generated from the Koyck model under the null hypothesis $\beta = 0$, that is, we set $\beta = 0$, $\mu = 0$, and we assumed variance 0.25, like in Table 1. For each simulation, the four test statistics were computed, and their values were stored. Finally, the four resulting samples were ordered in an ascending way, so that the quantiles became available. In each simulation, the sup absolute t and sup Wald statistics were obtained via a grid search over λ from 0 to 0.999 with step size 0.001.

Before we turn to an application of these tests to real-life data, it seems wise to see which of these tests performs best in practice. For that purpose, we ran another set of simulations, and the results are given in Table 3. Clearly, the power of the supremum tests is not very high, and hence we would recommend the use of the average tests.

3 An illustration

One might now be tempted to think (or hope) that the above considerations would not have a substantial impact on empirical analysis. Unfortunately, they do, as can already be illustrated for the illustrious monthly Lydia Pinkham data, which have been used in many marketing studies.

If one decides to neglect the MA part of the Koyck model, that is,

$$S_t = \mu + \beta A_t + \lambda S_{t-1} + v_t, \tag{6}$$

and replaces the intercept by twelve monthly dummies to correct for seasonal effects, then β gets estimated at a value of 0.360, with standard error 0.118, and λ is estimated equal to 0.370, with standard error 0.125.

Unrestricted estimation of the Koyck model, that is, (5) with monthly dummies, renders a β of 0.335 (0.101), a λ_1 of 0.690 with standard error 0.119, and a λ_2 of 0.561 with error 0.178. Clearly, these two λ parameters are rather different. Notice that the $t-$ratios for β in the above two models are 3.039 and 3.307, respectively.

The maximum likelihood estimates of β and λ for (2), again including monthly dummies, are found to be 0.339 (0.098) and 0.703 (0.125) respectively, where the reported standard errors are asymptotic standard errors. Consistent with the simulation evidence from Table 1, the estimates of λ_1 in the unrestricted model and λ in the restricted Koyck model are approximately equal (0.690 and 0.703). Furthermore, it can be seen that ignoring the MA part of the Koyck model indeed results in serious underestimation of the retention rate (0.370 versus 0.703), as suggested by Table 1. To put it differently, neglecting the MA component would result in a 90% duration interval of 1.3 months, whereas appropriate maximum likelihood estimation would result in a much longer 90% duration interval of 5.5 months[1].

The values of the four (average and supremum) statistics for testing $\beta = 0$ are 4.61, 22.06, 5.55 and 30.79, respectively. By comparing these realized values with the critical values in Table 2, we conclude that the two "ave tests" indicate a significant advertising effect at a 1% significance level, whereas the two "sup tests" fail to reject the null hypothesis at a 20% level. These contradictory results confirm our findings in Table 3 which suggests that the supremum tests do not have much power.

Finally, for illustrative purposes, Figure 1 shows the underlying distributions of the absolute t statistic $|t_\beta|$ and the Wald statistic t_β^2 over the different values of λ.

[1]Clarke (1976, p346) defines the $(100 \times \alpha)\%$ duration interval as the time period τ_α during which $(100 \times \alpha)\%$ of the expected cumulative advertising effect has taken place. It can be shown that $\tau_\alpha = \frac{\ln(1-\alpha)}{\ln(\lambda)} - 1$.

For both statistics, the supremum corresponds to $\lambda = 0.663$, which is quite close to the maximum likelihood estimate $\hat{\lambda} = 0.703$.

4 Conclusion

The Koyck model is often applied in marketing practice, but it is more complicated to analyze than one would perhaps think. Proper parameter estimation requires imposing parameter restrictions in the estimation routine. And, proper inference on the advertising effects requires new test statistics with non-standard asymptotic distributions, as the retention parameter disappears under the null hypothesis of no effect of advertising. In this paper we showed how these tools work. For the monthly Lydia Pinkham data we showed that various tests can lead to contrasting conclusions.

Table 1: Estimating the retention parameter λ in a Koyck model. In case (A) the MA term is neglected. In case (B) the model parameters are estimated using nonlinear least squares with unrestricted parameters λ_1 and λ_2, where only λ_1 is reported. The simulation results are based on 1000 replications. The cells contain the mean value of the 1000 estimates and the associated standard deviation. Advertising data are drawn from a standard normal distribution. Next, sales data are generated for $\mu = 0$, $\beta = 1$ and an error process having variance 0.25.

	n = 50		n = 500		n = 5000	
$\lambda = 0.5$						
A	0.412	(0.062)	0.420	(0.018)	0.420	(0.012)
B	0.502	(0.053)	0.500	(0.014)	0.500	(0.010)
$\lambda = 0.8$						
A	0.705	(0.053)	0.732	(0.013)	0.734	(0.004)
B	0.797	(0.025)	0.800	(0.005)	0.800	(0.002)
$\lambda = 0.9$						
A	0.825	(0.046)	0.855	(0.010)	0.859	(0.003)
B	0.899	(0.014)	0.900	(0.002)	0.900	(0.001)
$\lambda = 0.95$						
A	0.879	(0.051)	0.923	(0.008)	0.927	(0.002)
B	0.948	(0.012)	0.950	(0.008)	0.950	(0.000)

Table 2: Critical values of various tests for the hypothesis that $\beta = 0$ in the Koyck model. Number of replications is 40000. The sample size is 1000. The grid for λ runs from 0.000 to 0.999 with step size 0.001. Data are generated similar to those in Table 1.

Confidence level	Ave absolute t	Ave Wald	Sup absolute t	Sup Wald
80 %	1.21	1.67	6.18	38.13
90 %	1.53	2.56	7.47	55.86
95 %	1.80	3.50	8.41	70.65
99 %	2.34	5.79	9.80	95.99

Table 3: Empirical power of various tests for the hypothesis that $\beta = 0$ in the Koyck model. Number of replications is 1000 for each value of β. The sample size is 1000. The critical value is set at the 5% level. Data are generated similar to those in Table 1.

β	Ave absolute t	Ave Wald	Sup absolute t	Sup Wald
0.00	5.90	5.70	5.20	5.20
0.02	29.80	29.40	5.30	5.30
0.04	79.30	78.70	4.50	4.50
0.06	98.20	98.20	5.30	5.30
0.08	100.00	100.00	7.40	7.40
0.10	100.00	100.00	16.80	16.80

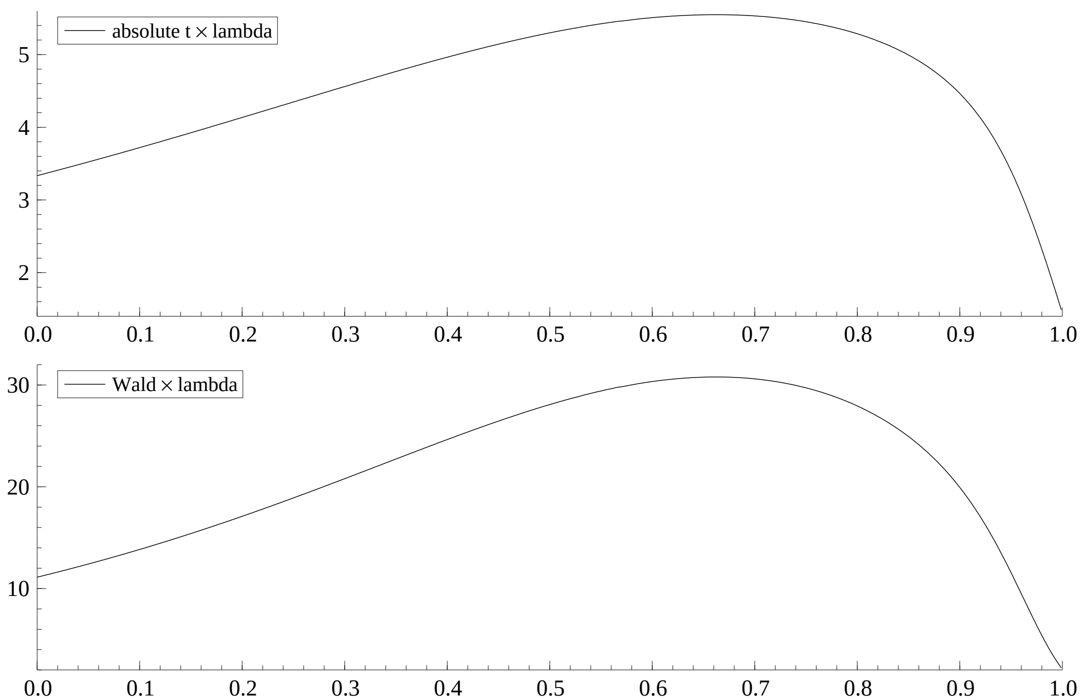

Figure 1: Testing for the significance of advertising effects in the Koyck model applied to the monthly Lydia Pinkham data. The values of the absolute t statistic and the Wald statistic are shown for different values of the retention parameter λ.

References

Andrews, D.W.K. and W. Ploberger (1994), Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, 62, 1383-1414.

Carrasco, M. (2002), Misspecified structural change, threshold, and Markov-switching Models, *Journal of Econometrics*, 109, 239 - 273.

Clarke D.G. (1976), Econometric measurement of the duration of advertising effect on sales, *Journal of Marketing Research*, 8 (November), 345 - 357.

Davies, R.B. (1987), Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, 64, 247-254.

Franses, P.H. (1991), Primary demand for beer in The Netherlands: An application of ARMAX model specification, *Journal of Marketing Research*, 28, 240-245.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press.

Hansen, B.E. (1996), Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, 64, 413-430.

Koyck, L.M. (1954), *Distributed Lags and Investment Analysis*, Amsterdam: North-Holland.