Paper 3841-2019

Using ETS® Software for Intervention Models for Sales of Cars

Anders Milhøj, University of Copenhagen

ABSTRACT

Cars in Denmark are heavily taxed, and the level of taxation is changed from time to time. In recent years, first an extra tax was introduced in order to reduce fuel usage, and two years later a new tax was introduced in order to make more secure cars affordable. The first intervention was in favor of small cars, while the second was in favor of large cars. Whether such changes in the taxation rate were introduced immediately or were announced some months before makes a difference in their impact. Also, changes in taxation are expected by everyone interested in the negotiations in parliament.

Statistics Denmark publishes the number of cars sold in many segments. In this paper, the car segments small, medium, and large are used in order to illustrate the effect of these changes in taxation. The models applied are estimated by the ARIMA procedure with intervention components, because exogenous, deterministic effects like changes in taxation should be modeled separately by deterministic components and not as a part of the stochastic model.

In this paper, models for exponentially decreasing impacts of an intervention are of interest. However, situations where impacts are increasing up to the actual date of an announced future change in taxation also exist in this data set. Such situations are easily modeled by reverting the direction of time.

INTRODUCTION

All time series models have a stochastic component to fit the influence from various sources outside the model. These impacts are in a statistical model seen as outcomes of stochastic variables. They are often denoted "noise" which is an intuitive explanation for the fact that they are seen as noise that makes the true signal, the correct data series, more dim. Often the series is however also influenced by events with large impacts on the time series. Such events are explicitly known as extraordinary.

We reserve the concept "interventions" to reactions to events that are of this non-stochastic nature. The existence of an exogenous intervention is known to the analyst and its impact should be modeled separately from the stochastic model of the "remaining" data series. The reaction to the time series of the intervention has some kind of deterministic explanation so the behavior of the series close to the time of the intervention has to be modeled apart from the stochastic mechanism that otherwise generate the observations. Otherwise the estimated residual variance become unrealistic large.

An example of such an event is the intake of medicine where biological measurements indicate a reaction to the intake. It is known that the medicine is given so any change in the physiological behavior of the patient is assumed to be due to the medicine. Another example is the reaction of the consumers to an increase in taxes, where the consumption is believed to decrease.

Intervention models are often formulated by dummy variables taken values like zero and one and hence they are by no means observed values. An intervention could model the level shift of a time series from summer to winter, but actual measurements of temperatures, humidity, hours of sunshine are by no way included in the model. For the two intuitive examples mentioned above this means that the actual dose of the intake and the actual level of the tax increase are not included in the model. More refined models for interventions are however very close to models relying on actual observed values of the dose or taxation percentages so the distinction is by no way strict.

A general framework for analyzing time series were presented in the book by Box and Jenkins(1976) and after that seasonal ARIMA models are standard when modelling univariate time series. Box and Jenkins(1976) in their book also considered intervention models of the kind considered in this paper.

TYPES OF INTERVENTIONS

Interventions may take various forms, as for instance one particular outlier, a level shift or perhaps a shift in the observed trend in the time series. More complicated interventions exist for instance a shift in the level which is not abrupt but instead occurs gradually.

An intervention for an intervention that only affect the series for just one observation, at time t_0, has the form

(1) $X_t = \omega I_t + Y_t$

where the intervention components I_t is given by $I_t = 0$ for all t but for just the value $I_{t0} = 1$ at the event of the event. More precisely $I_t = 0$ for $t \neq t_0$ and $I_t = 1$ for $t = t_0$. The remainder term Y_t could be modeled in many ways according to the situation, e.g. as a regression model or as a time series model such as a multiplicative, seasonal ARIMA model, see Box and Jenkins(1976).

For interventions for temporary level shifts we assumed above that the effect of the intervention was immediately, but often the effect lasts for some time. This could be modeled by introducing lags in the intervention model

(2) $X_t = \omega_0 I_t - \omega_1 I_{t-1} - .. - \omega_r I_{t-r} + Y_t$

Here the effect of the intervention is given as steps in the following way

Time t_0 : The effect is ω_0

Time t_0+1 : The effect is $-\omega_1$

..

Time t_0+r : The effect is $-\omega_r$

The choice of sign for the parameters ω_i in this parameterization is taken as the notation by Box & Jenkins(1976).

One problem by using $r > 0$ in this parameterization is that the number of ω−parameters could be large and as they tell the same story they are subject to multicollinearity. This means that the estimates of the ω_i parameters have large standard deviations.

If $r > 1$ a more *parsimonious* parameterization could be often more efficient. In order to reduce the number of parameters it is often better to adopt another form of parameterization using the notation of lag polynomials. The idea is that the effect is slowly decaying by a factor δ as

(3) $X_t = \omega I_t + \omega\delta I_{t-1} + .. + \omega\delta^r I_{t-r} + .. + Y_t = \frac{\omega}{1-\delta L} I_t + Yt$

Here the effect of the intervention is given as steps in the following way

Time t_0 : The effect is ω

Time t_0+1 : The effect is $\omega\delta$

..

Time t_0+r : The effect is $\omega\delta^r$

... etc.

The value $\delta = 0$ gives the simple intervention (1). But the situation of a positive parameter $0 < \delta < 1$ gives a gradual decrease towards the previous level of the series which is never reached but in practice the first steps are the largest. For $|\delta| \geq 1$ the series diverges and a new level is undefined. The parameter δ is therefore restricted to the interval $]-1, 1[$. Even if the value $\delta = 1$ is not included in the parameter space values of δ close to +1 takes the form of a step function. These kinds of models are discussed in detail by Box and Jenkins(1976).

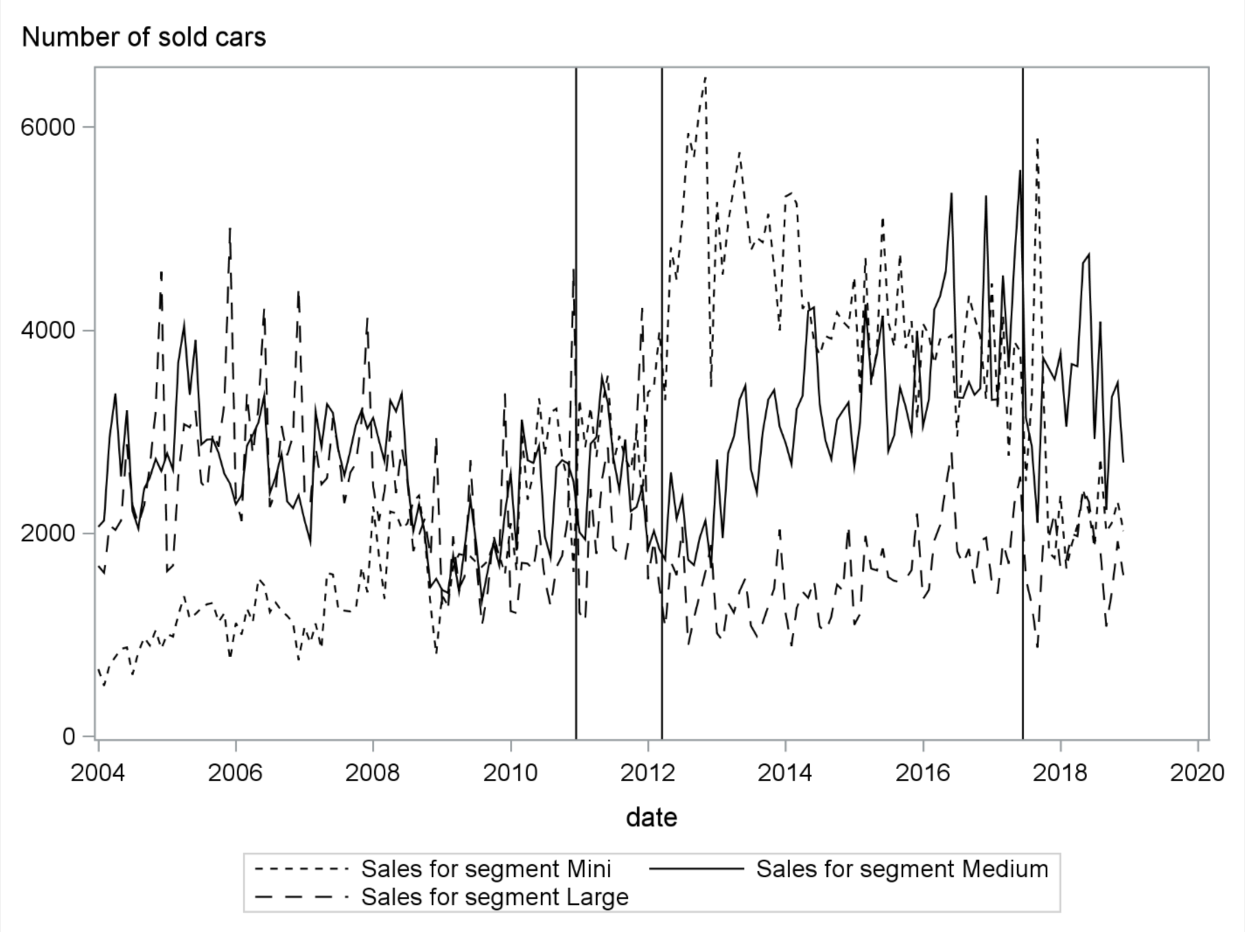

Figure 1 Danish sales of cars in three segments. The vertical lines represent major changes in taxation.

SALES OF NEW CARS IN DEMARK

Statistics Denmark publishes monthly data for the sales of new cars in form of many time series. In this paper we apply series for the sales numbers in three segments of cars for private use: Mini (like Volkswagen Up), medium (like Volkswagen Golf) and Large (like Volkswagen Passat). The time series are published with start January 2004 and data up to December 2018 is used in this paper - total of 180 observations.

The Danish taxation on new cars is heavy. For cheap cars a taxation is at present 85% is added to value, while the taxation of the values higher than a certain limit, around 185,000 DKK (corresponding to approximately 30,000 USD), is as high as 150%. Previously these percentages were even higher 105% resp. 180% and the limit was lower than now, around 85.000 DKK (corresponding to approximately 13,000 USD). On top of the actual price of the car and all specific taxes moreover the general value added tax adds 25% to the final price.

By definition this taxation system favors sales of smaller, cheaper cars. But in recent years the system has changed somewhat, as a new tax in 2013 depending on the fuel economy was introduced, and in 2017 a new taxation (in fact a tax reduction) depending on security factors was introduced. While taxes depending on fuel economy favor smaller cars as they use less fuel, taxes depending on security favors large cars as it is safer to be a passenger in a large car in case of a frontal crash.

The sales series for the three segments series are plotted in Figure 1. Figure 2 shows average prices as met by the consumers in Danish Kroner for the three segments. The vertical lines on both plots mark some of the changes in taxation as an indicator of large interactions in the sales.

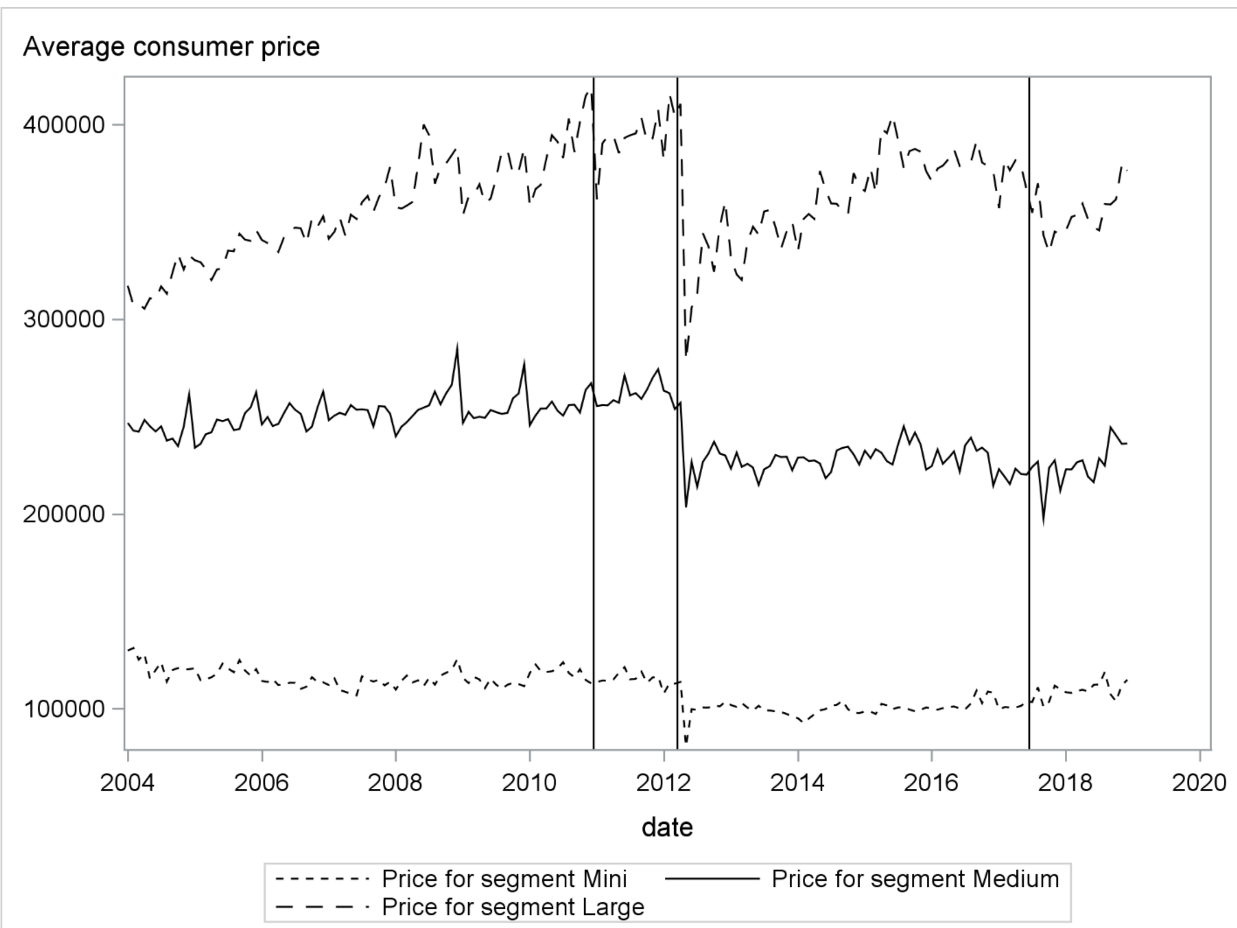

Figure 2 Average price of cars in three segments. The vertical lines represent major changes in taxation.

ARIMA MODELS FOR MONTHLY SALES OF CARS

As a benchmark model the so called airline model is used. This model is often applied as a default model for seasonal time series; made famous by the example of the number of airline passengers by Box and Jenkins (1976). The model is also applied as default in the X11-ARIMA algorithm for seasonal adjustments. For these datasets the model with parameters estimated by maximum likelihood are

Mini: $(1 - B)(1 - B^{12})X_t = (1 - 0.51B)(1 - 0.88B^{12})\varepsilon_t$, $var(\varepsilon_t) = 532.76^2$

Medium: $(1 - B)(1 - B^{12})X_t = (1 - 0.59B)(1 - 0.71B^{12})\varepsilon_t$, $var(\varepsilon_t) = 462.51^2$

Large : $(1 - B)(1 - B^{12})X_t = (1 - 0.66B)(1 - 0.60B^{12})\varepsilon_t$, $var(\varepsilon_t) = 400.55^2$

AUTOMATIC OUTLIER DETECTION

The variances are heavily influenced by extreme observations as they are estimated using the sum of squared residuals. Two times the standard deviation is of course a much too narrow limit for outlier detection as the expected number of outliers in that case is $0.05 \times 180 = 9$. The critical value for a 5% test for outliers among 180 observations is 3.92 times the standard deviation. However, the residual variance is rather large due to the existence of outliers, which are nor corrected for any interventions. So a limit of three times the residual standard deviation in the ARIMA models are applied.

For the three segments we find some months where the residuals have absolute values larger than three times their standard deviation, but only one observation for each segment is larger than four times its standard deviation. For the segment mini for September 2017 which was 2796 higher than expected because of political negotiations on a possible increased taxation on smaller cars. For the segment medium a similar outlier was found in December 2016 where the taxation on all cars were reduced. For the segment large an outlier is found for December 2010 were sales were 1413 higher than expected due to an increased taxation on cars with low fuel economy.

If these residuals are removed using dummy variables, the residual variance is reduced significantly. Of course more outliers could be compensated by dummy variables in order to reduce the variance even further, but the number of parameters then of course increases as each outlier leads to an extra parameter.

INTERVENTIONS FOLLOWED BY AN EXPONENTIAL DECAY

A numerically large residual could be considered as the simplest form of an intervention which is easy to see as a part of an ordinary time series analysis. Interventions followed by an exponential decay (3) are more advanced and not so simple to identify. But by macro programming it is possible to fit an intervention of the form (3) for each month in the dataset. However, the denominator variable δ is not well defined if the intervention is applied for the very last part of the series. Table 1 presents the estimated values of ω and δ for the months were ω numerically exceeds four times its standard deviation and the t-value (estimated value divided by its standard deviation) exceed three. In some situations, the estimated δ is close to one, which means the intervention is probably a step function. Several of the entries in table show values of δ not significantly different from 0.7, which are interventions of the form (3).

Mini

Date	ω	stderr ω	δ	stderr δ
JUL2012	1666.29	392.89	0.96	0.05
AUG2012	1343.83	430.23	0.92	0.09
NOV2012	1460.84	386.92	-0.47	0.18
DEC2012	-1924.00	453.04	0.62	0.18
SEP2017	3016.34	401.60	0.22	0.13
NOV2017	-1988.33	408.34	0.99	0.03

Medium

Date	ω	stderr ω	δ	stderr δ
MAR2012	-1035.25	336.03	0.90	0.10
DEC2016	1741.20	359.15	-0.09	0.20

Large

Date	ω	stderr ω	δ	stderr δ
NOV2008	-802.31	259.45	0.93	0.08
NOV2011	1208.11	306.63	0.61	0.17
MAR2013	-948.96	239.31	0.99	0.02

Table 1 Estimated parameters of models for exponential decay

OUTLIER DETECTION IN THE REVERTED SERIES

This model is fitted in order to take care of the autocorrelation structure in the time series. The estimated ARMA parameters and the residual variance has the same values when they are estimated based on the reverted time series. However, the idea of an innovation process, ε_t, and hence also the residual values are different. The residuals are now prediction errors when backcasting one period using only future values of the time series. In this case we also see outliers, but they are not necessarily for the same time stamps as outliers in the usual forecasts.

For the three segments we find a few months where the residuals have absolute values larger than three times their standard deviation and again only one for each segment are larger than four times the standard deviation. That is the sales in the segment mini for September 2017 which was 2797 higher than expected, for the segment medium the sales December 2016 were 1324 larger than expected, and the sales for the segment large for December 2010 where the sales were 1413 larger than expected.

EXPONENTIAL EFFECTS BEFORE AN INTERVENTION

If the reverted series are the in the model for exponential decay, we find interventions which is in actual time direction exponentially build up in the months before rather than decaying after the interventions. The following table gives the estimated values of ω and δ for the months were ω numerically exceeds four times its standard deviation.

For some of these interventions the estimated δ is close to zero. This means that the intervention this particular month is probably an outlier. For one month, the estimated δ is close to one. This means the intervention is probably a step function. Several of the remaining entries in Table 2 show values of δ not significantly different from say 0.7; that is of the form (3).

These results are based on an estimation of the denominator parameter δ for each month in the dataset. In order to ease the estimation burden a fixed value of δ, say δ = 0.7, could be applied so that only the ω parameter has to be estimated.

Mini

Date	ω	stderr ω	δ	stderr δ
NOV2012	2336.31	412.41	0.80	0.09
DEC2012	-1540.25	404.76	-0.29	0.22
SEP2017	2819.17	399.02	0.06	0.22
NOV2017	-1381.68	447.43	0.03	0.32

Medium

Date	ω	stderr ω	δ	stderr δ
DEC2016	1625.74	360.00	-0.07	0.21
SET2017	-1315.20	382.75	0.46	0.25
AUG2018	1275.06	392.00	-0.09	0.30

Large

Date	ω	stderr ω	δ	stderr δ
JUN2006	961.77	319.77	0.51	0.26
DEC2010	1463.87	310.96	0.05	0.21
MAR2011	1192.94	307.22	0.64	0.16

Table 2 Estimated parameters for exponential decay models for the reverted series

STATISTICS CANNOT REPLACE KNOWLEDGE!

In the analyses presented for the three series interventions are detected automatically by a screening process. In order to see whether these findings are due to well defined events like e.g. changes in the taxation, we have to look through all history books and newspapers in the past many years for information of the history of Danish car taxes - and what else could affect car sales. Hmmm - sounds cumbersome, even though Google is your friend.

It is optimal if a statistical model includes all known external factors for car sales. A more honest analysis takes the approach first to identify all possible reasons for interventions with an ideas of the form of the intervention and after that estimate the parameters of the intervention components. Again - sounds cumbersome, even though Google is still your friend.

If the purpose of the analysis is mainly focused on the most recent part of the time series a screening, like the one presented above is, however, a practical approach to help the analysts to remember long forgotten events.

REFERENCES

Box, G.E.P. and Jenkins, G.M. (1976), *Time Series Analysis: Forecasting and Control*, Revised Edition, San Francisco, CA: Holden−Day.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Anders Milhøj
Department of Economics
University of Copenhagen
anders.milhoj@econ.ku.dk