Pre-adjustment process of real retail trade series in Croatia

Josip Arnerić, Anita Čeh Časni
University of Zagreb, Faculty of Economics and Business Zagreb, Croatia

Ante Rozga
University of Split, Faculty of Economics Split, Croatia

Keywords
Calendar effects, time series, pre-adjustment process, country specific regressors.

Abstract

Economic activity fluctuations are often influenced by various factors related to the calendar. These factors include non-working (non-trading) days, leap years, public holidays etc. In analysis of many economic variables in which time series are seasonally adjusted it is necessary to identify and correct calendar effects using suitable method, but there is no general or unique procedure for correcting these effects in pre-adjustment process. If these effects are not well adjusted, the identification of the ARIMA model may not be correct, and the quality of the seasonal adjustment is poor. Therefore, in these paper different methods of correction of the calendar effects are compared and applied to time series of real retail trade turnover (RRT) in Croatia (monthly data observed from January 2001 to December 2013). The most common method used is regression model with different types of explanatory variables which take into account calendar effects. The contribution of this paper is to define new explanatory variables (regressors that include not only different number of working and non-working days of the month but also country specific calendar effects) which will give most accurate correction of RRTT time series.

1. Introduction

Short-term statistics are often characterized by seasonal fluctuations and other calendar and trading-day effects, which can camouflage relevant short and long-term movements of the series, and obstruct a clear understanding of analyzed economic phenomena. Therefore, the main aim of seasonal adjustment is to remove changes that are due to seasonal or calendar influences to produce a clearer picture of the underlying behavior of the analyzed series according to Cleveland and Devlin (1980).

Even particular adjustment methods are similar; for example method that uses U.S. Census Bureau compared to the Eurostat method, it is not clear which method is better. The choice of method usually depends on experience and the policy of entire organization such as Croatian Bureau of Statistics. Calendar adjustment as well as seasonal adjustment is a source of constant debate, due to the different methods that can be used, and the different tools and computer programs that exist.

According to INE (2013), seasonal fluctuations are movements that occur with a similar intensity each month (quarter) or each season of the year, and which are expected to continue occurring. Furthermore, they define the calendar effect as the impact produced in the time series, due to the different structure that the months (or quarters) present in the different years (in both length and composition), even if the remaining factors influencing the variable of interest remain constant.

Most of economic series are observed on a monthly or quarterly basis, but months (quarters) are not comparable due to different number of working and non-working days (not the same number of Mondays, Tuesdays, etc.). This may have an impact on the observed variables such as retail sales, industrial productions and transportation. For example, retail sale turnover is likely to be more important on Saturdays than on other days of the week. Hence, there is a need for revision of the data adjusted for seasonal effects and calendar effects with two main reasons: they are a result of the revision of the gross data, due to an improvement in the information (in terms of coverage and/or reliability) or they are a result of a better estimation of the seasonality pattern, due to new information provided by new gross data, and due to the characteristics of the filters and procedures that eliminate the calendar and seasonal components by Palma and Marini (2004).

Even so, despite being a frequent practice, seasonal adjustment is still a source of constant debate resulting from different methods applied, different tools and computer programs that exist. Moreover, the manipulation of the original data that might occur through seasonal adjustment is also questioned. According to Arteche et al. (2011), calendar effects can be divided into two groups. The first group includes the effects of working (or trading) days and the second group deals with special calendar effects, such as Christmas, Easter or other (national) holidays. During the data processing, the abovementioned effects have to be taken into account and various methods of seasonal adjustment related to them must be applied. An inappropriate or poor quality of seasonal and calendar adjustment can generate false signals and can negatively affect the interpretation of the adjusted data. Having all this in mind, the aim of this paper is to apply different methods of correction of the calendar effects of real retail trade turnover (RRTT) in Croatia and define new explanatory variables which will give most accurate correction of RRTT time series. Therefore, new regressors of calendar effects adjustment will be defined.

The reminder of this paper is as follows. After the introduction in section 2 the different methods and alternative models for calendar effect correction are presented. Section 3 will give the results of the empirical analysis. Finally, in section 4 concluding remarks will be given.

2. Overview of different methods of calendar effects correction

There are many different methods for correction of calendar effects. The empirical analysis in this paper is focused on the linear regression method with different regressors for calendar adjustment of time series. Namely, corrections of working (or trading) days are carried out from the estimation of the linear regression. So, in a manner of Arteche et al. (2011) the following model of calendar structure using the explanatory variables is defined:

$$y_t = z_t'\gamma + \varepsilon_t$$
$$\varphi(L)\delta(L)\varepsilon_t = \theta(L)v_t, \qquad t = 1,2,\ldots,T$$

where y_t is the time series of interest, ε_t are error terms that follow an ARIMA process, $\varphi(L)$, $\delta(L)$ and $\theta(L)$ are finite polynomials of the lag operator L, z_t' is a vector of $(K\times 1)$ of K relevant regressors, γ is a $(K\times 1)$ vector of unknown parameters and v_t is an error term defined as white noise.

Component $z_t'\gamma$ represents nonstochastic effects which are subtracted from the original series before applying the ARIMA methodology for decomposition of the time series into trend/cycle, seasonal and irregular component. The simplest nonstochastic effect is the average, in this case the regression constant. More complex effects are for instance the intervention variables, atypical observations or calendar effects. Each time period is characterized by different number of Mondays, Tuesdays, …, Sundays, therefore economic activity (in our case the real retail trade turnover) can be affected by this fact.

The working days effect differentiates working days from weekend days, so, in relevant literature, variable (we_t) is usually used to express the weighted difference between the number of working days (w_t) and non-working days (nw_t) during the analyzed period *t*. This is defined as:

$$we_t = \left(w_t - \frac{5}{2} nw_t \right)$$

where the number of non-working days is multiplied by 5/2 so that the average of the newly created variable was zero. The coefficient of the variable $\boldsymbol{we_t}$ includes the effect of additional working days in the period *t*.

It must be taken into consideration that the length and composition of each month or quarter has a seasonal part which has to be captured in the seasonal component, and must not be eliminated. For example, March always has 31 days, and on average, it has more Mondays than February. Moreover, the working-day adjustment must only be associated with the non-seasonal part of the effect. Namely, the non-seasonal part of the composition of the working days of the month (quarter) may be estimated by the deviation of this number of days from its long-term average. In order not to eliminate seasonality, the regressors must be calculated in deviations from the average for each month/quarter.

According to INE (2013), in order to eliminate the seasonal part of this effect the following working day regressor has to be calculated:

$$(w_{(m/t)T} - \bar{w}_{m/t}) - \frac{\bar{w}}{\bar{n}} \cdot (n_{(m/t)T} - \bar{n}_{m/t})$$

where:

$w_{(m/t)T}$ is the number of working days in month (m) or quarter (t) from year T,

$\bar{w}_{m/t}$ is the average of the number of working days for each month m = January, February,... December, or quarter = I, II, III, IV, calculated over a 28-year calendar (in shorter series, this average of working days each month or quarter is obtained with the calendar of the complete series, and is recalculated every year),

$n_{(m/t)T}$ is the number of non-working days each month (m) or quarter (t) in year T,

$\bar{n}_{m/t}$ is the average of the number of non-working days in each month m = January, February, ... December, or quarter = I, II, III or IV, calculated over a 28-year calendar (in shorter series this average is recalculated every year),

$\frac{\bar{w}}{\bar{n}}$ is the quotient between the average of the number of working days and the average of the non-working days, calculated over a 28-year calendar. In shorter series, this quotient is obtained with the calendar of the complete series, and is recalculated every year; though it may stabilize near the same value in long series.

According to equation two the quotient between the average number of the working days and the average number of the non-working days equals 5/2 and it is assumed to be constant over time.

Furthermore, Eurostat and the European Central Bank emphasize the importance of leap year correction, which can be modelled using the following zero mean variable:

$$lpy_t = \begin{cases} 0.75 & \text{if } t = \text{February of leap year} \\ -0.25 & \text{if } t = \text{February of non-leap year} \end{cases}$$

= 0 if t= other months

The adjustment to moving holidays aims to eliminate those values that are affected by events following a complex pattern over the years from the series. In this paper, Easter is moving holiday that most affects analyzed series. This effect is partially seasonal, since on average, it is celebrated more often in April than in March. Given that the seasonal part must be captured in the seasonal component, it must not also be eliminated with the correction of the Easter holiday effect. Namely, the regressor(s) for the Easter holiday must be built in such a way that they intend to capture the effect which this moving holiday can have on the economic series of interest. It is not possible to specify a standard way of preparing this regressor, since the repercussion that this holiday may have is very different from one series to another. For more details on this regressor please refere to INE (2013).

When building the regressor for working days, Croatian national holidays have to be taken into account. In the calculation of both working and non-working days, this matter considers the number of each of these days in each month in Croatia weighted by the specific weight that Croatia has in the economic series that is being adjusted for this effect (in our case the real retail trade turnover) .

3. Alternative models for the calendar effect correction

Since the aim of this paper is to apply different methods of correction of the calendar effects of real retail trade turnover (RRTT) in Croatia and to define new explanatory variables which will give most accurate correction of RRTT time series, in this section five models which allow for different corrections of the calendar effects will be presented. Namely, the model called A1 is the basic alternative since it is highly similar to the model which some European NSI use for the correction of the calendar effect. Other alternative models generalize A1 by adding new variables for other effects (A2 and A3) or simplify it (A4 and A5) performing goodness of fit of the model and its prediction power. In table 1, abovementioned alternative models are given.

However, alternative A2 is most popular model which combines standard regression analysis with ARIMA modeling before seasonally adjustment of the series. This model is known as Reg-ARIMA model and it is a part of X-12-ARIMA technique.

Model		Variable description
A1	$y_t = \beta_0 + \beta_1 PHC_t + \beta_2 we_t + \beta_3 lpy_t + \varepsilon_t$ $PHC_t = PH_t - DF$	we_t-defined in (2) lpy_t - defined in (4) PH_t-the number of non-working days corresponding to Monday, Tuesday, ..., Friday in the month t DF- the long run average of non-working days in the month excluding weekends
A2	$y_t = \beta_0 + \beta_1 PHC_t + \sum_{i=1}^{6} \beta_{i+2} w_t^i + \varepsilon_t$ $w_t^1 = (Mon_t - Sun_t), \ldots, w_t^6 = (Sat_t - Sun_t)$	-
A3	$y_t = \beta_0 + \beta_1 PHC_t + \beta_2 lpy_t + \sum_{i=1}^{6} \beta_{i+3} w_t^i + \varepsilon_t$	-
A4	$y_t = \beta_0 + \beta_1 WDSP_t + \varepsilon_t$ $WDSP_t = WD_t - PH_t - DL$	WD_t- the number of working days PH_t-the number of non-working days corresponding to Monday, Tuesday, ..., Friday $WDSP_t$- the effect of working days, trading days, Easter effect and implicitly the effect of leap year DL-the average number of working days in period t
A5	$y_t = \beta_0 + \beta_1 CE_t + \varepsilon_t$ $CE_t = (WD_t - PH_t) - RDF(We_t + PH_t)$	We_t- the number of Saturdays and Sundays during the period t, similar to the variable we_t defined in (2) but it includes the effect of non-working days which are not weekends RDF- the ratio between working days and non-working days

Table 1: Alternative models for calendar effect correction

Model A2 analyzes the effect of trading days. Model A3 includes the leap year correction with the variable lpy_t defined by Eurostat.

According to abovementioned alternatives the one which is a parsimonious (model with one regressor) and includes effect of working days, Easter effect and implicitly the effect of leap year is alternative A5. Therefore, alternative A5 will be applied to real retail trade turnover series with modification according to equation (3). Proposed modification is based on two following assumptions:

- The ratio between average number of working days and average number of non-working days *RDF* is not constant (it should be recalculated every year).
- Saturdays and Sundays are working days of the week.

Furthermore, proposed modification of alternative A5 according to equation (3) will be compared to model defined in equation (2). The empirical results are presented in section 3.

It is rational to assume that Saturdays and Sundays are working days of the week due to retail sales in Croatia. Working on weekends, especially on Sundays became an important marketing element and usual practice in Croatia, in accordance with arrival of foreign retail chains and the formation of new national retail chains which mainly trade in nutrition products and consumer goods. Large shopping centers are becoming a destination for family trip on Sundays and skilled traders constantly organize actions and promotions in order to attract more customers.

4. Empirical analysis

The data set analyzed in this paper is series of real retail trade turnover indices in Croatia. The sample period is from January 2001 to December 2013 for a total of 156 monthly observations. According to parasimony principle two regression models are estimated using different single regressor which includes all calendar effects.

First regression model uses regressor as defined in previous equation. Results of this regression model are obtained with OLS method (Figure 1).

	Coefficient	Std. Error	t-ratio	p-value
const	97.4660	9.99898	9.7476	0.0000
REG_1	0.127767	0.375777	0.3400	0.7343

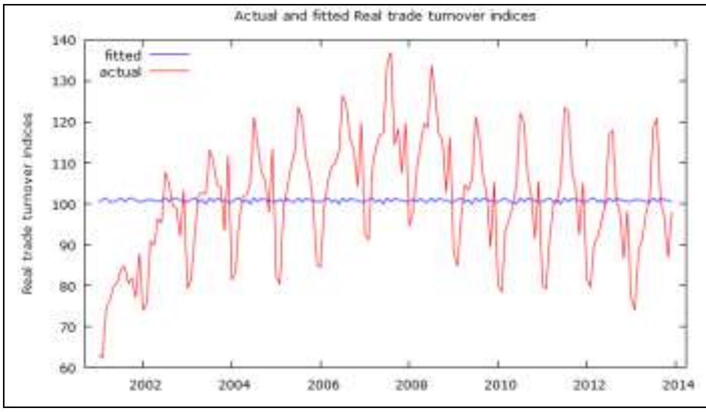
Mean dependent var	100.8416	S.D. dependent var	14.81265
Sum squared resid	33983.73	S.E. of regression	14.85509
R^2	0.000750	Adjusted R^2	-0.005739
$F(1, 154)$	0.115604	P-value(F)	0.734315
Log-likelihood	−641.2893	Akaike criterion	1286.579
Schwarz criterion	1292.678	Hannan–Quinn	1289.056
$\hat{\rho}$	0.695241	Durbin–Watson	0.568444

Figure 1: Estimated OLS regression model with REG_1 as regressor

According to figure 1 effect of additional working days (estimated coefficient of REG_1) have positive sign which means that on average retail trade turnover series should be upward corrected for 0,1277 when taking calendar effects into account.

Non-significance of this coefficient is expected due small variations that can be explained by calendar effects (Figure 3). However, the significance tests as well as other regression diagnostics are not as important as subtraction fitted values from original series to compute calendar adjusted values.

Therefore, calendar adjusted values are residuals plus 100 which is exactly the mean of retail trade turnover indices. Due to lack of space actual values (RTI), fitted values and residuals only for year 2002 are presented as a part of Figure 2, while actual values and calendar adjusted values are presented on Figure 3.



	RTI	fitted	residual
2002:07	107.768171	101.426763	6.341408
2002:08	104.990402	100.532397	4.458005
2002:09	99.497267	101.298996	-1.801729
2002:10	99.035243	101.426763	-2.391520
2002:11	92.101212	100.851813	-8.750601
2002:12	103.381128	100.532397	2.848731
2003:01	79.299137	100.532397	-21.233260
2003:02	81.553753	101.043463	-19.489711
2003:03	92.723625	101.426763	-8.703137
2003:04	102.249251	100.404630	1.844620
2003:05	102.739777	100.979580	1.760198
2003:06	102.443868	99.957448	2.486420
2003:07	113.246120	101.426763	11.819357

Figure 2: Actual values (RTI), fitted values and residuals

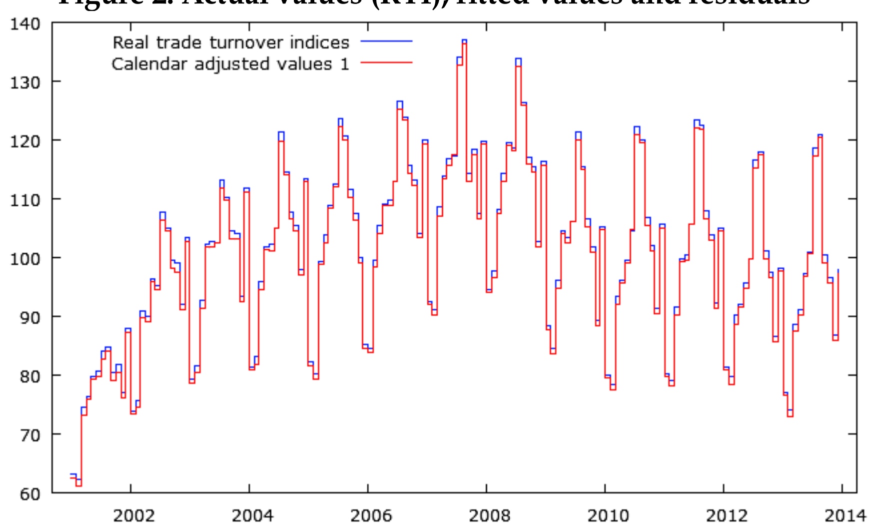


Figure 3: Actual values of RTI and calendar adjusted values

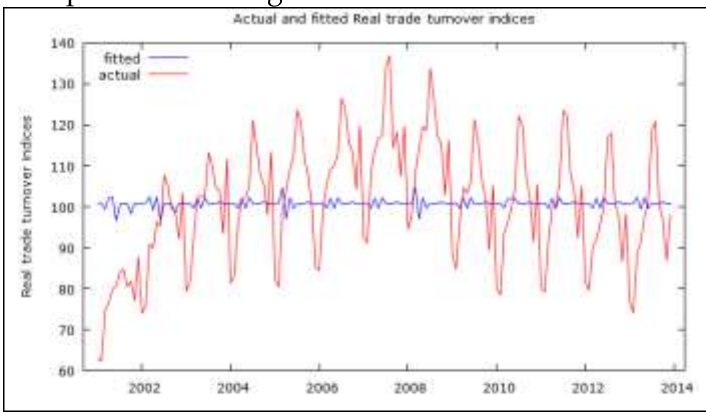
Second regression model uses newly proposed regressor defined as modification of alternative A5 according to equation (3). Results of this regression model are obtained with OLS method (Figure 4).

	Coefficient	Std. Error	t-ratio	p-value
const	100.852	1.18666	84.9885	0.0000
REG_2	−0.0960931	0.105280	−0.9127	0.3628

Mean dependent var	100.8416	S.D. dependent var	14.81265
Sum squared resid	33826.25	S.E. of regression	14.82063
R^2	0.005381	Adjusted R^2	-0.001078
$F(1, 154)$	0.833094	P-value(F)	0.362806
Log-likelihood	−640.9271	Akaike criterion	1285.854
Schwarz criterion	1291.954	Hannan–Quinn	1288.332
$\hat{\rho}$	0.686898	Durbin–Watson	0.584245

Figure 4: Estimated OLS regression model with REG_2 as regressor

According to figure 4, effect of additional working days (estimated coefficient of REG_2) have negative sign which means that on average retail trade turnover series should be downward corrected for -0,096 when taking calendar effects into account. Non-significance of this coefficient is expected due small variations that can be explained by calendar effects (Figure 6). Calendar adjusted values are residuals plus 100 which is exactly the mean of retail trade turnover indices. Due to lack of space actual values (RTI), fitted values and residuals only for year 2002 are presented as a part of Figure 5, while actual values and calendar adjusted values are presented on Figure 6.



	RTI	fitted	residual
2002:07	107.768171	100.852334	6.915837
2002:08	104.990402	100.852334	4.138068
2002:09	99.497267	100.852334	-1.355067
2002:10	99.035243	98.379167	0.656075
2002:11	92.101212	100.852334	-8.751121
2002:12	103.381128	100.852334	2.528794
2003:01	79.299137	100.852334	-21.553197
2003:02	81.553753	100.874509	-19.320757
2003:03	92.723625	99.696048	-6.972423
2003:04	102.249251	102.008619	0.240631
2003:05	102.739777	99.696048	3.043729
2003:06	102.443868	102.201334	0.242534
2003:07	113.246120	100.852334	12.393786

Figure 5: Actual values (RTI), fitted values and residuals

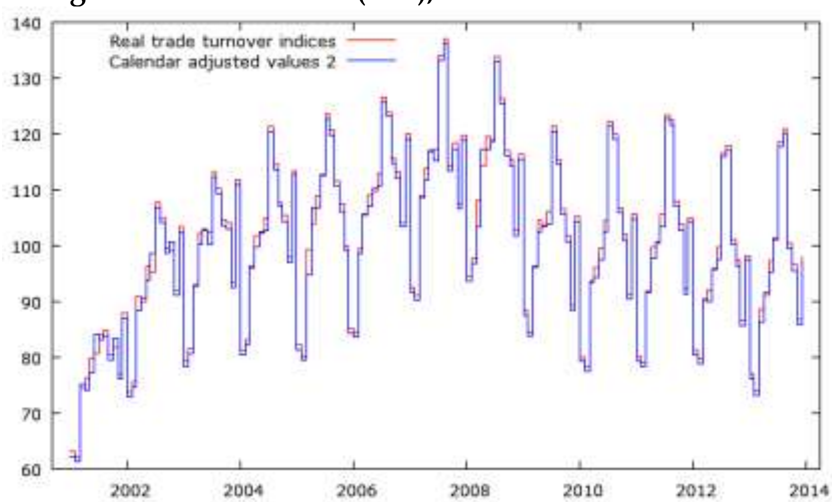


Figure 6: Actual values of RTI and calendar adjusted values

In figure 7, two series of calendar adjusted values are compared with original series of retail trade turnover indices in Croatia (due to lack of space only results for 2001 – 2006 are presented). The first series of calendar adjusted values (CA_1) are obtained according to residuals of first regression model with REG_1 regressor.

The second series of calendar adjusted values (CA_2) are obtained according to residuals of second regression model with REG_2 regressor. Regressor REG_2 is newly proposed as modification of alternative A5.

	RTI	CA_1	CA_2		RTI	CA_1	CA_2
2001:01	63.1351	62.6027	62.2828	2004:01	81.3510	80.8186	80.4986
2001:02	62.2251	61.1816	61.3506	2004:02	83.1131	81.9418	82.3347
2001:03	74.6614	73.2347	75.1581	2004:03	96.0352	94.6085	96.3424
2001:04	76.4270	76.0223	74.2256	2004:04	101.8355	101.4309	99.8237
2001:05	79.7629	79.2305	77.3367	2004:05	102.2026	101.2231	102.5098
2001:06	80.6731	79.8213	84.0926	2004:06	104.9192	104.9618	102.7142
2001:07	84.0766	82.6499	83.2243	2004:07	121.2922	119.8655	120.4399
2001:08	84.7291	84.1967	83.8767	2004:08	114.6109	114.0785	113.7585
2001:09	80.5255	79.2265	79.6732	2004:09	107.8561	106.5571	107.0038
2001:10	81.8557	80.4289	83.4765	2004:10	105.5800	104.6004	104.3412
2001:11	76.9766	76.1248	76.1242	2004:11	97.9442	97.0924	97.0919
2001:12	87.9200	87.3876	87.0676	2004:12	113.5095	112.9771	112.6572
2002:01	73.9098	73.3774	73.0575	2005:01	82.2320	81.6996	81.3797
2002:02	75.6907	74.6472	74.8162	2005:02	80.3231	79.2796	79.4485
2002:03	90.8276	89.8480	88.4014	2005:03	99.4126	98.8802	94.7060
2002:04	89.9667	89.1149	90.6882	2005:04	103.8913	102.5923	106.8933
2002:05	96.3502	95.8178	93.9240	2005:05	108.9574	108.4250	106.7561
2002:06	95.3419	94.4901	98.7614	2005:06	112.4816	112.0770	112.7856
2002:07	107.7682	106.3414	106.9158	2005:07	123.6417	122.2149	122.7894
2002:08	104.9904	104.4580	104.1381	2005:08	120.6314	120.0990	119.7791
2002:09	99.4973	98.1983	98.6449	2005:09	111.6006	110.3016	110.7483
2002:10	99.0352	97.6085	100.6561	2005:10	107.4156	106.4360	106.1778
2002:11	92.1012	91.2494	91.2489	2005:11	100.0000	99.1482	99.1477
2002:12	103.3811	102.8487	102.5288	2005:12	85.1689	84.6365	84.3165
2003:01	79.2991	78.7667	78.4468	2006:01	84.5081	83.9757	83.6557
2003:02	81.5538	80.5103	80.6792	2006:02	99.4860	98.4426	98.6115
2003:03	92.7236	91.2969	93.0276	2006:03	105.4332	104.0064	105.7371
2003:04	102.2493	101.8446	100.2406	2006:04	109.1777	108.7730	107.1691
2003:05	102.7398	101.7602	103.0437	2006:05	109.8385	108.8589	110.1424
2003:06	102.4439	102.4864	100.2425	2006:06	112.9222	112.9647	110.7208
2003:07	113.2461	111.8194	112.3938	2006:07	126.5786	125.1518	125.7262
2003:08	110.2213	109.6889	109.3690	2006:08	123.9354	123.4030	123.0831
2003:09	104.5821	103.2831	103.7298	2006:09	115.6388	114.3398	114.7864
2003:10	104.1513	103.1717	102.9135	2006:10	113.2893	112.3097	112.0515
2003:11	93.3357	92.4839	92.4833	2006:11	104.1850	103.3332	103.3327
2003:12	111.7542	111.2218	110.9019	2006:12	119.8972	119.3648	119.0449

Table 2: Two series of calendar adjusted values (CA_1 and CA_2) compared with original series of retail trade turnover indices for first six years of observation

For example, significant deviation between calendar adjusted values 1 and 2 is evident for October 2002. At this month the value of original series of retail trade turnover index is 99,03; the calendar adjusted value 1 is 97,61; the calendar adjusted value 2 is 100,65; and the difference between calendar adjusted values is 3,1%.

This difference can be explained that in October 2002 there is no non-working days (public holidays), the number of working days in October are maximal (31 days), which included Saturdays and Sundays (CA_2).

5. Conclusion

Economic activity fluctuations are often influenced by various factors related to the calendar. These factors include non-working (non-trading) days, leap years, public holidays etc. In analysis of many economic variables in which time series are seasonally adjusted it is necessary to identify and correct calendar effects using suitable method, but there is no general or unique procedure for correcting these effects in pre-adjustment process. In this paper different methods of correction of the calendar effects are compared and applied to time series of real retail trade turnover (RRT) in Croatia (monthly data observed from January 2001 to December 2013). The most common method used is regression model with different types of explanatory variables which take into account calendar effects. The contribution of this paper is to define new explanatory variables (regressors that include not only different number of working and non-working days of the month but also country specific calendar effects) which will give most accurate correction of RRTT time series.

Estimation results show that newly proposed regressor REG_2, as modification of alternative A5 according to equation (3) is better alternative for calendar adjustment of retail trade turnover series. This can be justified by taking into account the two important assumptions:

(a) the ratio between average number of working days and average number of non-working days *RDF* is not constant (it should be recalculated every year) and

(b) Saturdays and Sundays are working days of the week.

This analysis also indicates that the detection and correction of the calendar effect should consider other factors which are not included in the models of the five analyzed alternatives.

References

Arteche J., Majovská R., Mariel P. and Orbe S. (2011): Detection and Correction of Calendar Effects: An Application To Industrial Production Index Of Álava, *Journal of Applied Mathematics*, 4 (3).

Cleveland W., Devlin, S. (1980): Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods, *Journal of the American Statistical Association*, 75, pp. 487-496.

Di palma, f., marini, m.: *the working/trading day adjustment of italian quarterly National accounts: methodology and presentation of the main result*. In proceeding of joint Unece/eurostat/oecd meeting on national accounts, geneva, ces/ac.68/2004/12, 2004.

Eurostat: *follow-up of the cmfb task force on seasonal adjustment of quarterly national Accounts*. Eurostat unit b2, eurostat b1-b2/cn 514, 2002a.

Eurostat: *methodology of short-term business statistics- interpretation and guidelines.* Eurostat commission, theme 4, industry trade and services, 2002b.

Hungarian Central Statistical Office (2007). Seasonal Adjustment Methods and Practices, European Commission Grant 10300.2005.021-2005.709, Budapest, July 2007, Final Version 3.1, contributors: Foldesi, E., Bauer, P., Horvath, B., Urr, B.

INE (2013). INE Standard for the Adjustment of Seasonal Effects and Calendar Effects in Short-term Series, National Statistics Institute, March 2013.