Modelling Wages in Croatia Using a Second Order Polynomial Regression Model

Tea Baldigara[1] & Kristina Duvnjak[1]

[1] Faculty of Tourism and Hospitality Management, Opatija, Croatia

Correspondence: Tea Baldigara, Faculty of Tourism and Hospitality Management, Primorska 42, 51410 Opatija, Croatia. Tel: 385-51-294-684. E-mail: teab@fthm.hr

Abstract

As known, wages are a significant factor of the economic activity. They affect personal consumption, the living standard and the GDP. Due to the importance of wages for the national economy the paper aimed to analyse annual average wages in Croatia in the last two decades, namely from 1994 to 2013. For these purposes a second order polynomial regression model is designed. The empirical results have shown that, the chosen model performed well in terms of MAPE and, presented a reasonably good predictability performance. All performed diagnostic statistics showed that the specified models passed all the tests and that they fit the data reasonably well throughout the sample period.

Keywords: annual average wages, Croatia, modelling, time-series model, second order polynomial regression model

1. Introduction

In the early 90-ties real wages in Croatia experienced a drastic fall. In 1992 they accounted for 43.5% of Croatian GDP. Due to changes in the definition and the structure nominal wages in 1995 experienced a significant upward trend. From 1995 to 2000 the cumulative growth of nominal net wages accounted for 83,1% of the GDP with an average annual growth rate of 8.9%, while productivity increased on average for 4.3% per year. 1992 and 1993 were years of seriously low real wages due to high inflation and high ascending unemployment rate (from 16.1% in 1995 to 21.5% in 2000). In 1999 Croatian GDP registered a decrease of 0.4%, the unemployment an increase of 11.9%, but on the other side there was an increase of 10.1% in real wages. In 2013 the average unemployment growth rate was 20.3% (as in January 2015) and it ascended for 1.2 percentage points compared with 2012. The main issue of Croatian economy is that economic growth does not lead to a satisfactory decline of the unemployment rate. In fact the growth of real wages occurs contemporaneously with the growth of the unemployment rate and a significant growth of the labour costs. The crises that affected Croatian economy in 2008 caused a decrease in GDP of 6%. The downward trend in GDP strongly influenced the structure and the level of wages as well as the global living standard. Although, there is a slight upward trend in Croatian wages in the last few decades, there is also a significant upward trend in costs of living and therefore a decrease in personal consumption. The latest data from the Croatian Bureau of Statistics reveals that the average monthly net wage in Croatia is 5.716 Croatian Kuna (745 euros) that is well below the average in many developed countries. The average monthly paid off net earning per person in paid employment in legal entities in the Republic of Croatia was 2.9% (nominal) and 3.4% (real) higher in December 2014, than it was compared to December 2013. Data show that the highest average net salaries were currently being paid in the air transport sector, where average monthly salaries were 10.555 Kuna (1.375 euros), jobs in marketing and research were next where workers on average took earned 9.353 Kuna a month (1.220 euros), with those in the financial sector just behind with an average of 9.039 (1.175 euros) while computer programmers and consultants were close behind with 9.001 Kuna or 1.170 euros. (http://www.tradingeconomics.com/croatia/wages). Jobs with an average of less than 4.000 Kuna (520 euros) were those in the service industry, wood manufacturing, administration and furniture retail. (http://www.croatiaweek.com/average-monthly-wage-in-croatia-a-miserly-e720/)

2. Data and Methodology

The paper analyses the average real wages in Croatia from 1994 to 2013. Data are annual and are taken from the National Bureau of Statistics.
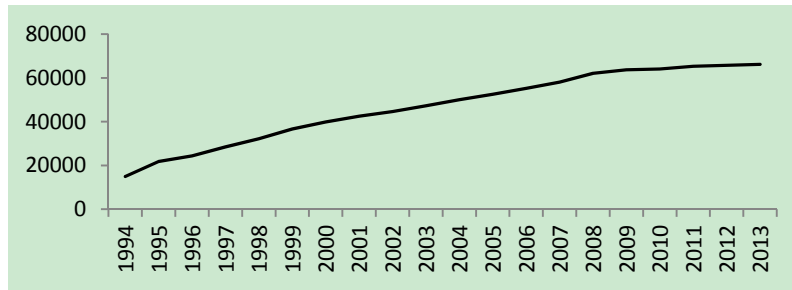
Figure 1. Annual average real wages (in mil Croatian Kuna) in Croatia from 1994 to 2013

Figure 1 clearly shows a nonlinear upward trend of wages in Croatia in the analysed period. In the attempt to model the data generating process for the empirical time series-data a second order polynomial regression function is used. A polynomial regression is a special case of multiple regression, in which the relationship between the independent variable and the dependent variables is modelled as an nth order polynomial. It fits a nonlinear relationship between the observed variables, but as a statistical estimation problem it is linear, because the regression function is linear in the unknown regression parameters. The goal of regression analysis is to model the expected value of a dependent variable y in terms of the value of an independent variable (or vector of independent variables) x. In general, the expected value of y as an nth order polynomial, yielding the general polynomial regression model can be expressed as it follows:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n + \varepsilon \tag{1}$$

where n is some integer.

Polynomial regression is a particular form of a multiple linear regression. A polynomial expression in (1) is an equation that is linear in parameters having an independent variable that is raised to a power included in the model. When using polynomial regression models, the classical assumptions for regression analysis should be applied. In order to obtain valid estimates of the parameters, the following assumptions have to be made (Song, Witt, Li, 2012):

- The expected value of the dependent variable depends on the values of the explanatory variables and the unknown β parameters. This is equivalent to $E(e_{it}) = 0$.

- The sample variance of the dependent variable or the variance of the error term remains constant over time. If this assumption does not hold, the model suffers from the problem of heteroscedasticity.

- Any two observations on the dependent variable or the residuals are not correlated. Violation of this assumption results in a model with autocorrelated residuals.

- The values of the residuals are normally distributed about their mean: $e_{it} = N(0, \sigma^2)$.

- The values of the explanatory variables are known and there is no linear relationship among the explanatory variables. If this condition is not true, the model suffers from the problem of multicollinearity.

It is assumed that the empirical time-series data of the wages in Croatia in the period from 1994 to 2013, in dependence of time t (the independent variable x) can be approximated by a second-order polynomial (quadratic) function as follows:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \tag{2}$$

where

     Y   = the dependent variable

     β_0 = intercept constant

     β_1 = linear effect parameter

     β_2 = quadratic effect parameter

     X   = time, explanatory variable

     e   = residual

The model in (2) is said to be hierarchical because it contains all terms of order two and one. Only hierarchical models are invariant under linear transformation. Polynomial models can be useful in situations where curvilinear

effects are present in the response function and in approximating functions to unknown and possible very complex nonlinear relationships. All lower order terms must be in the equation in order for the polynomial function to be represented, but the statistical tests should focus on the second order term, as the highest term in the model. In essence, the tests are performed to establish whether the highest term in the model is significantly different from zero and thus whether the polynomial fits the data. There are also arguments from physics which suggest that first-degree and second-degree polynomials in t, which in fact are linear and quadratic time trends are common in the natural word and therefore such trends might also arise in the social world.

3. Empirical Results and Discussion

The second-order polynomial regression model in (2) is fitted using the method of ordinary least squares and the results for the estimated quadratic polynomial trend model are as follows:

$$W= 11627.3+ 4561.76X- 88.695X^2 \tag{3}$$

$$t = 1 \text{ in } 1994$$

$$\text{unit for } t= 1 \text{ year}$$

$$\text{unit for } Y = \text{million HR Kuna}$$

$R^2 adj$=0.995   RSS=18788759 $\chi^2 Auto$(1)=4.15999 $\chi^2 Norm = 2.57$

$\chi^2 White$(4)=4.185095   VIF=17.705

where

- W are the average wages in million Croatian Kuna
- x is the predictor, time
- RSS is the sum squared residuals
- χ^2_{Norm} is the Jarque-Bera normality test
- $\chi^2_{Auto}(1)$ is the Breusch-Godfrey test for autocorrelation
- $\chi^2_{White}(4)$ is the White test for heteroscedasticity
- VIF is the variance inflation factor

As shown in (3) the value of the variation inflation factor shows high collinearity between the predictors in the estimated model. As it is known, collinearity is a common problem in polynomial regressions, where the predictors are usually highly correlated. A general rule states that, if the VIF of a variable exceeds 10, the variable is considered to be highly collinear. The variance inflation factor estimates how much the variance of a coefficient is ''inflated'' because of linear dependence with other predictors. The minimum possible value of the VIF is 1. The problem of multicollinearity can be removed or reduced substantially by standardizing the linear and the quadratic terms in the polynomial regression equation in (3). Thus, independent variable X is standardized using the following linear transformation:

$$S = (X - \bar{X})/S_X \tag{4}$$

where

- S is the standardized variable X
- $\bar{X}$ is the sample mean of X
- S_X is the sample standard deviation

Likewise, the quadratic term is created by taking squared values of the standardized linear term in (3). One of the advantages of standardization is that multicollinearity among linear and quadratic terms is substantially reduced (Table 1)

Table 1. VIF values before and after standardizing

```
Variance Inflation Factors

Minimum possible value = 1.0
Values > 10.0 may indicate a collinearity problem

              x    17.705
             x2    17.705

VIF(j) = 1/(1 - R(j)^2), where R(j) is the multiple correlation coefficient
between variable j and the other independent variables

Properties of matrix X'X:

 1-norm = 769636
 Determinant = 2.334948e+008
 Reciprocal condition number = 1.9515411e-006

Variance Inflation Factors

Minimum possible value = 1.0
Values > 10.0 may indicate a collinearity problem

              s    1.000
             S2    1.000

VIF(j) = 1/(1 - R(j)^2), where R(j) is the multiple correlation coefficient
between variable j and the other independent variables

Properties of matrix X'X:

 1-norm = 51.381429
 Determinant = 5445.9429
 Reciprocal condition number = 0.10856932
```

The correlation coefficient declined significantly too (Table 2).

Table 2. The correlation matrix between predictors before and after standardizing

```
corr(x, x2) = 0.97134820                 corr(s, S2) = 0.00000000
Under the null hypothesis of no correlation:   Under the null hypothesis of no correlation:
 t(18) = 17.3402, with two-tailed p-value 0.0000   t(18) = 0, with two-tailed p-value 1.0000
```

Moreover, the equation (3) and (5) provide the same fit and the same results in the same value of adjusted R-square (compare model (3) and model (5)). After standardizing the model is re-estimated and the results are shown below:

$$W = 49746.9 + 15968.5S - 3104.32S^2 \tag{5}$$

t = 1 in 1994

unit for X = 1 standard deviation

unit for Y = million HR Kuna

R^2adj=0.995   RSS=18788759 $\chi^2 Auto(1)$=4.15999 $\chi^2 Norm = 2.57$

$\chi^2 White(4)$=4.185095   VIF=1

where

- W are the average wages in million Croatian Kuna
- S and S^2 are the standardized terms of the predictors
- RSS is the sum squared residuals
- χ^2_{Norm} is the Jarque-Bera normality test
- $\chi^2_{Auto}(1)$ is the Breusch-Godfrey test for autocorrelation
- $\chi^2_{White}(4)$ is the White test for heteroscedasticity
- VIF is the variance inflation factor

Furthermore, the performed diagnostic statistics shows that the coefficients of the standardized variables are significant at 5% level. The computed JB-statistics is 2.57 and is smaller than the critical value of $\chi^2(2) = 5.99$, and therefore the hypothesis of normally distributed residuals for the selected model can be accepted. To test the model for the presence of heteroscedasticity the White test is used. The null hypothesis of no heteroscedasticity can be accepted as the calculated White statistics, of 4.185095 is smaller than the critical value of $\chi^2(4) = 9.48$. The presence of serial correlation is tested using the Breusch-Godfrey test. As the calculated value $\chi^2(1) = 4.1599$ exceeds the critical value of 3.84146 the null hypothesis of no serial first order correlation cannot be accepted. It is well known "....that under both heteroscedasticity and autocorrelation the usual OLS estimators, although linear, unbiased and asymptotically normally distributed, are no longer minimum variance among all linear unbiased estimators; they may not be best linear unbiased estimators (BLUE). As a results, the usual, t, F and χ^2 may not be valid." (Gujarati, 413). Autocorrelation can be reduced by applying the generalized least square (GLS) method which uses the differencing technique in order to obtain non autocorrelated variables. GLS is equivalent to applying ordinary least squares to a linearly transformed version of empirical data. Knowing the consequences of autocorrelation the model (5) has to be transformed. Assuming that the error term in (2) follows the AR(1) scheme, namely,

$$u_t = \rho u_{t-1} + \varepsilon_t \quad -1 < \rho < 1 \tag{6}$$

the generalized equation can be used:

$$Y_t^* = \beta_0 \cdot (1 - \rho) + \beta_1 X_{1t}^* + v_t \tag{7}$$

where

- $v_t = u_t - \rho \cdot u_{t-1}$
- $Y_t^* = Y_t - \rho \cdot Y_{t-1}$
- $X_{1t}^* = X_{1t} - \rho \cdot X_{t-1}$

Therefore, ρ is estimated from the residuals and then the residuals $\hat{u}_t$ are regressed on $\hat{u}_{t-1}$. The following regression is run:

$$\hat{u}_t = \rho \cdot \hat{u}_{t-1} + v_t \tag{8}$$

where

$\hat{u}_t$ – are the residuals obtained from the original model regression

v_t – are the error term of this regression

The obtained regression is:

$$\hat{u}_t = \rho \cdot \hat{u}_{t-1} + v_t \tag{9}$$

Using the OLS method on equation (8) the following results are obtained:

$$\hat{u}_t = 0.461518 \cdot \hat{u}_{t-1} + v_t \tag{10}$$

Once ρ has been estimated, the quasi-differenced variable $x_t - \rho x_{t-1}$, for every explanatory and for the dependent variable, are computed. Using these estimates the standardised model as per equation (5) is transformed and estimated. The results are as follows:

$$W = 26779.1 + 15553.1 SCO - 2909.69 S2CO \tag{11}$$

t = 1 in 1995

unit for X= 1 standard deviation

unit for Y = million HR Kuna

$R^2 adj$=0.985 RSS= 13129466 $\chi^2 Auto(1)$=1.2766 $\chi^2 Norm = 4.987$

$\chi^2 White(4)$=4.15 VIF=1.31

where

- *W* are the average wages in million Croatian Kuna
- *SCO* and *S2CO* are the predictors
- *RSS* is the sum squared residuals

- χ^2_{Norm} is the Jarque-Bera normality test
- $\chi^2_{Auto}(1)$ is the Breusch-Godfrey test f or autocorrelation
- $\chi^2_{White}(4)$ is the White test for heteroscedasticity
- VIF is the variance inflation factor

After transformation, the model satisfies all, the usual assumptions of the classical linear regression model and all the re-performed performance tests showed that the model fit the data well. As shown in Figure 2 the model fit the data reasonably well throughout the sample period and has the same predictive capability;
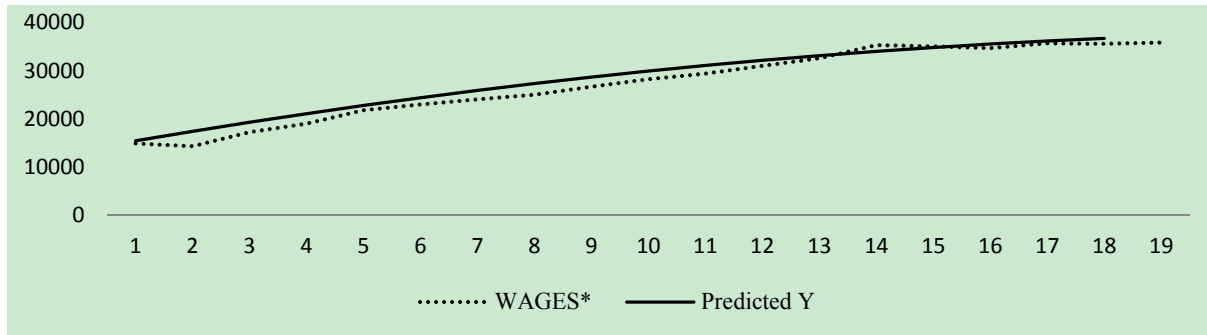

Figure 2. Wages in Croatia: fitted versus actual; Time bounds: 1995-2013

The models fit the general movement of the analysed series during the entire sample period. The forecasted results are presented in the table below.

Table 3. Forecasted results of the estimated models

	German tourist arrivals
$R^2 adj$	0.9850
MAPE	2.6904
RMSE	831.2800
Theil'U	0.40369

The forecast results are reasonably good; the predicted values, in fact, are quite close to the actual values.

4. Conclusion

As known, wages are a significant factor of the economic activity. They affect personal consumption, the living standard and the GDP. Due to the importance of wages for the national economy the paper aimed to analyse annual average wages in Croatia in the last two decades, namely from 1994 to 2013. Data shown that average wages in Croatia increased throughout the analysed period. In the attempt to model the average annual wages a second order polynomial regression model was used. All performed diagnostic statistics showed that the specified models passed all the tests and that they fit the data reasonably well throughout the sample period. The empirical findings could be considered as starting points of future detailed and more systematic quantitative analysis of wages in Croatia. In order to enable future empirical researches of wages in Croatia and collecting valuable pieces of information, a comprehensive desk research and review of modelling as well as forecasting literature was carried out. In reviewing the literature there has been noted a lack of systematic and detailed domestic quantitative researches focused on analysing the patterns and core determinants of wages in Croatian economy. In future, more sophisticated quantitative, both extrapolative and causal forecasting techniques, should be applied to investigate, model and forecast wages in Croatia. Profoundly aware that modelling and forecasting are challenging and controversial issues, that the adequacy of a forecasting model is valued according to its out-of- sample forecasts, and that is still difficult to indicate which model or class of models is more adequate for tourism demand modelling, the authors of this study

wished to highlight the necessity of more systematic quantitative analysis of wages in Croatia's economy in all its determinants.

References

Baggio, R., & Klobas, J. (2011). *Quantitative Methods in Tourism*. Bristol: Channel View Publications.

Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Croatian Employment Service National Coordination Office for EURES. (2014). *Living and Working in Croatia*. Zagreb, Croatia.

Dobre, I., & Alexandru, A.A. (2008). Modelling unemployment rate using Box-Jenkins procedure. *Journal of applied quantitative methods, 3*(2), 156-166.

Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. New York: John Wiley&Sons

Ekonomski institut Zagreb. (2014) *Analiza plaća u javnom i privatnom sektoru u Hrvatskoj*. Retrieved from http://www.eizg.hr/hr-HR/Analiza-placa-u-javnom-i-privatnom-sektoru-u-Hrvatskoj-1229.aspx

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. Columbus:McGraw-Hill.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. New York: Wiley and Sons.

Nestić, D., Lovrinčević, Ž., & Mikulić, D. (2001). Plaće u Hrvatskoj- stanje i makroekonomske implikacije različitih scenarija budućih kretanja. *Economic Review, 52*(1-2), 234-277.

Neter, J. et al. (2004). *Applied Linear Statistical Models*. Coumbus: McGrae-Hill/Irwin

Parkin, M. (1990). *Macroeconomics*. Ontario: Addison-Wesley Publishing Company, Ontario.

Song, H., Witt, S., & Gang, L. (2012). *The Advanced Econometrics of Tourism Demand*. London: Routledge.

Weisang, G., & Awazu, Y. (2008). Vagaries of the Euro: an Introduction to ARIMA Modelling, *CS-BIGS, 2*(1), 44-45. Retrieved January 4, 2012, from http://www.bentley.edu/csibgs/vol2-1/jaggia.pdf