THE APPLICATION OF SARIMA AND VAR MODELS IN TIME SERIES OF CZECH CONSTRUCTION AND COMPARISON RESULTS

Ondřej Šimpach

Abstract

The purpose of this study is provide to stakeholders simple, but relevant quarterly estimates of the average number of persons employed in the Czech construction and the average gross monthly wage per actual person in the Czech construction for 2012. These statistics are very important because we can incorporate them as variables to adaptive expectations. The construction is a sector with great importance to the national economy. For 2012 the Czech market does not expect significant interferences from the government and the international perspectives expect stagnation. Therefore, for an initial analysis it will be used the methodology by Box and Jenkins (1970) and then using the time series of gross domestic product at constant prices it will be explored the Granger causality (Granger, 1969). After the confirmation of the existence of a causal relationship, the VAR models will be constructed, and based on them will be provided more precise quarterly estimates of the average number of persons employed in the Czech construction and the average gross monthly wage per actual person in the Czech construction for 2012. With this particular example will be shown, that even simple predictions based on the stochastic trend can provide simple, fast and cheap vision about the development of some important indicators of the national economy.

Keywords: construction of Czech Republic, Box-Jenkins, Granger causality, SARIMA, VAR

1 INTRODUCTION

In a small open economy with a growing and dynamic industry, more and more come into the forefront requests for estimates of important national economic indicators and performance indicators of the economy. These requirements include claims on the speed and precision however we know that both claims can not be applicable simultaneously. It is confirmed that in the short term that is not affected by unexpected changes in the behavior of government and mass changes in the preferences of private entities, it is possible to obtain sufficient point estimates, which are not differ from the estimates from leading institutions. These sufficient point estimates can be obtained quickly and cheaply for a period of one year, especially when the market expects stagnation. The more conservative trend to the future is expected, these simple point estimates provide more precise results. The paper will analyze the quarterly time series of "average number of persons employed in the Czech construction" and "average gross wages per actual person in the Czech construction" by methodological approach of the authors Box and Jenkins (1970), based on this the simple forecasts of the average number of persons employed and the average gross wages for the period of 2012 will be counted. Using quarterly time series "Gross domestic product at constant prices" will be explored the Granger causality (Granger, 1969), whether the trend contained in this series may affect the series of average number of persons employed in the Czech construction, or a series of average gross wages per actual person in the Czech construction. If it will find the proof about existence of causality, will be constructed vector autoregression models, which provide more accurate predictions of the average number of employed and average gross wages per actual person for the period of 2012. The forecasts from both approaches will be graphically compared and commented. It is expected, that the outputs of these two approaches will not be too deferent,

and despite their simplicity the economic entities operating in the Czech market will used them as relevant estimates for their adaptive expectations.

2 STOCHASTIC MODELING – THE AVERAGE NUMBER OF PERSONS EMPLOYED

The first part will deal with analysis of quarterly time series of average number of persons employed in the Czech construction, which will certainly contains seasonality. The statistics publishes the Czech Statistical Office at regular intervals and observations are known from 1st quarter of 2000 to 3rd quarter of 2011. Based on the methodological approach of Box and Jenkins (1970) there were estimated the SARIMA model parameters running EViews and the output of this process is represented in Tab. 1.

Tab. 1 – The estimated model parameters, Average number of persons employed.
Source: Eviews, own construction.

Dependent Variable: AV_No_of_PERS_EMP				
	Coefficient	*Std. Error*	*t-Statistic*	*Prob.*
C	388782.0	53468.63	7.271216	0.0000
AR(1)	0.820235	0.090448	9.068588	0.0000
SAR(4)	0.928569	0.065869	14.09712	0.0000
R-squared	0.865353	Mean dependent var	395877.6	
Adjusted R-squared	0.858448	S.D. dependent var	11677.20	
S.E. of regression	4393.353	Akaike info criterion	19.68232	
Sum squared resid	7.53E+08	Schwarz criterion	19.80644	
Log likelihood	-410.3288	Hannan-Quinn criter.	19.72782	
F-statistic	125.3235	Durbin-Watson stat	1.790278	
Prob(F-statistic)	0.000000			

The estimated model has the form *SARIMA (1, 0, 0) (1, 0, 0)c*. Due to recommendations that refer e.g. Hylleberg et al. (1990), the time series did not have to be differentiated before the start of analysis, because at the 5% significance level showed stationarity. The model can be expressed as

$$AV\_No\_of\_PERS\_EMP = 388782.0 + [AR(1) = 0.820235, SAR(4) = 0.928569]$$

All model parameters are the 5% significance level statistically significant, the correspondence between model and data expressed by $R^2 = 0.865$ is sufficient and the corrected version by the degrees of freedom $\overline{R^2} = 0{,}858$ is also high. Using Breusch and Godfrey autocorrelation test, referred e.g. by Box and Pierce (1970) there was not rejected the null hypothesis about absence of autocorrelation at the 5% significance level. The ARCH test, referred e.g. by Engle (1995) did not reject the null hypothesis testing homoskedasticity at the 5% significance level, so it is possible to claim that the model is estimated correctly as a whole.

Now it is possible to calculate the forecast for the 4th quarter of 2011 by the estimated model above, because this value will be officially published by the CZSO after the presentation of this article and next one it is possible to calculate the forecast for subsequent 4 quarters of 2012. The time series for these predictions has sufficient length and in the national economy is expected the quiet development all year 2012. It is important to note, that in case of unexpected intervention to the economy by external force inducted e.g. by government or political strategies from other countries which should have the strong impact on the market in the Czech Republic, all these preconditions losing the effect. The published seasonally adjusted time series of average number of persons employed in construction by CZSO from 1Q2000 to 3Q2011 is displayed in Fig. 1 in red, the estimated forecast for 4Q2011 to 4Q2012

is displayed in blue. Better preview of trend will be further illustrated by graphic output with using seasonally adjusted series by X12 – ARIMA method.

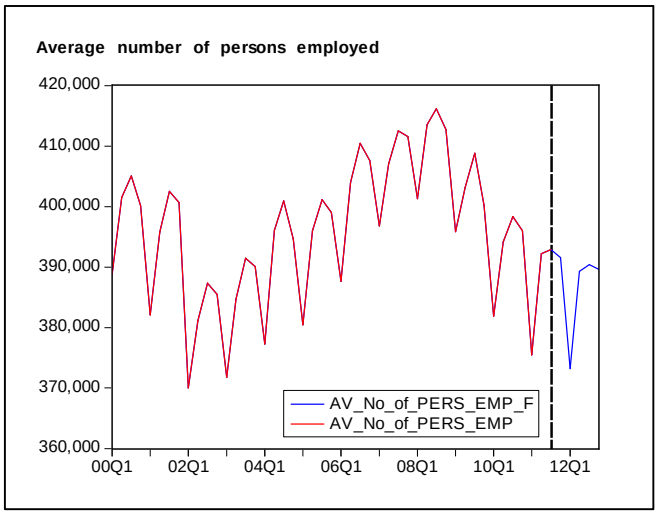


Fig. 1 – Forecast of average number of persons employed, not seasonally adjusted.
Source: CZSO - data, own construction – forecast.

For the purpose of vector autoregression model there was obtained the quarterly time series of gross domestic product at constant prices from 1st quarter of 2000 to 3rd quarter of 2011. Given that the null hypothesis of absence causality, referred e.g. by Granger (1969) was at the 5% significance level rejected (see Tab. 2), it could be constructed VAR model with one lag.

Tab. 2 – Granger causality. Source: EViews, own construction.

VAR Granger Causality/Block Exogeneity Wald Tests			
Dependent variable: GDP_SA			
Excluded	*Chi-sq*	*df*	*Prob.*
AV_No_of_PERS_EMP_SA	5.099643	1	0.0493
All	5.099643	1	0.0493
Dependent variable: AV_No_of_PERS_EMP_SA			
Excluded	*Chi-sq*	*df*	*Prob.*
GDP_SA	5.087085	1	0.0479
All	5.087085	1	0.0479

The estimated parameters of VAR(1) model are shown in Tab. 3. Using Breusch Godfrey autocorrelation test, there was not rejected the null hypothesis of absence autocorrelation at the 5% significance level and using White heteroscedasticity test, referred e.g. by Hamilton (1994), there was not rejected the null hypothesis at 5% significance level.

Tab. 3 – The estimated parameters of VAR(1) model. Source: EViews, own construction.

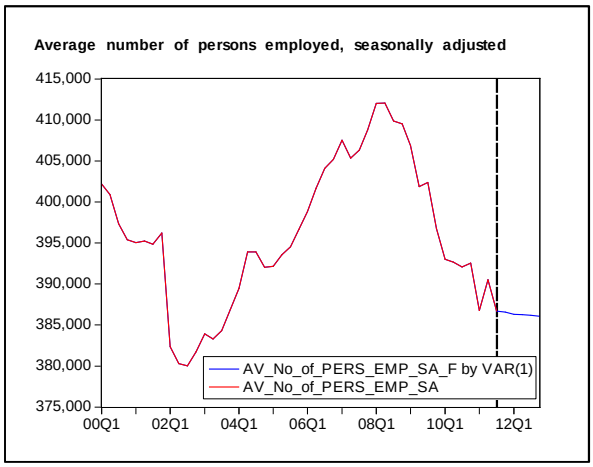
Vector Autoregression Estimates		
Standard errors in ( ) & t-statistics in [ ]		
	GDP_SA	*AV_No_of_PERS_EMP_SA*
GDP_SA(-1)	0.983254	0.001535
	(0.01336)	(0.00520)
	[73.6106]	[0.29510]
AV_No_of_PERS_EMP_SA(-1)	0.048340	0.996082
	(0.02661)	(0.01037)
	[1.81649]	[96.0962]
R-squared	0.953245	0.871475
Adj. R-squared	0.943092	0.868553
Sum sq. resids	3.01E+09	4.56E+08
S.E. equation	8266.489	3219.853
F-statistic	6469.787	298.3446
Log likelihood	-479.1672	-435.7950
Akaike AIC	20.92031	19.03456
Schwarz SC	20.99982	19.11407
Mean dependent	789165.7	395808.7
S.D. dependent	99456.15	8880.993

The VAR(1) model can be expressed as

$$GDP\_SA_t = 0.983254\ GDP\_SA_{t-1} + 0.048340\ AV\_No\_of\_PERS\_EMP\_SA_{t-1}$$

$$AV\_No\_of\_PERS\_EMP\_SA_t = 0.001535\ GDP\_SA_{t-1} + 0.996082\ AV\_No\_of\_PERS\_EMP\_SA_{t-1}$$

Seasonally adjusted time series of average number of persons employed is shown in Fig. 2 and Fig. 3 in red. In blue there are shown the forecasts for 4Q2011 – 4Q2012. The estimated values by VAR(1) model are shown in the left graph, the estimated values by SARIMA (1, 0, 0) (1, 0, 0)c model are shown in the right graph.



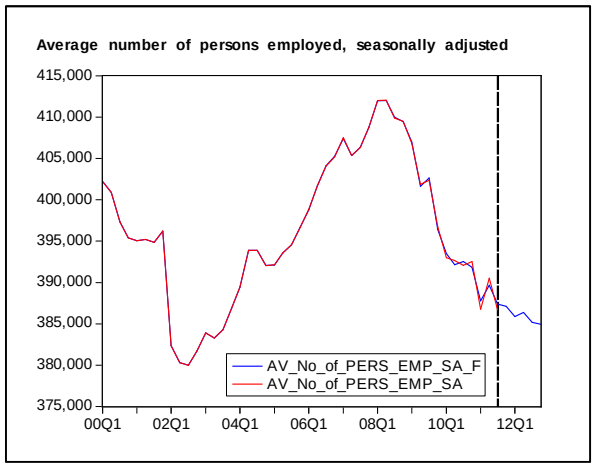


Fig. 2 (left), Fig. 3 (right) – Forecast of average number of persons employed (left – by VAR model, right – by SARIMA model), seasonally adjusted.
Source: CZSO – data, own construction – forecast.

It is evident, that the estimates of trend for 2012 are slightly more pessimistic by the SARIMA model. The estimates by the VAR model are quieter. However, both models expecting pessimistic trend, i.e. a decrease of average number of persons employed. The

concrete values for each quarter are shown in Tab. 4 for the VAR model and in Tab. 5 for the SARIMA model.

Tab. 4 (left), Tab. 5 (right) – The estimated values of average number of persons employed. Source: own construction.

OBS.	*AV_No_of_PERS_EMP_SA_F by VAR(1)*
4Q2011	386 564.6
1Q2012	386 288.5
2Q2012	386 244.2
3Q2012	386 176.6
4Q2012	386 047.0

OBS.	*AV_No_of_PERS_EMP_SA_F*
4Q2011	387 098.3
1Q2012	385 879.9
2Q2012	386 391.1
3Q2012	385 188.3
4Q2012	384 956.1

3 STOCHASTIC MODELING – THE AVERAGE GROSS MONTH WAGE PER ACTUAL PERSON

Second part of the analysis refers to the quarterly time series of average gross monthly wage per actual person in the Czech construction, which as well as time series of average number of persons employed contains seasonality. The statistics publishes the Czech Statistical Office at regular intervals and observations are known from 1st quarter of 2000 to 3rd quarter of 2011. By EViews system there were estimated SARIMA model parameters and the output of this is represented in Tab. 6.

Tab. 6 - The estimated model parameters, Average gross monthly wage per actual person. Source: Eviews, own construction.

Dependent Variable: D(AV_GR_MONTH_WAGE)				
	Coefficient	*Std. Error*	*t-Statistic*	*Prob.*
AR(4)	1.010423	0.048726	20.73684	0.0000
MA(1)	-0.544233	0.133663	-4.071687	0.0002
R-squared	0.945887	Mean dependent var	254.0915	
Adjusted R-squared	0.944535	S.D. dependent var	1971.674	
S.E. of regression	464.3504	Akaike info criterion	15.16560	
Sum squared resid	8624852.	Schwarz criterion	15.24835	
Log likelihood	-316.4777	Hannan-Quinn criter.	15.19593	
Durbin-Watson stat	2.129583			

It is important to note that the time series of average wages was non-seasonal differentiated before the analysis, due to recommendations that refer e.g. Hylleberg et al. (1990). The time series showed non-stationarity at the 5% significance level. The estimated model has form *SARIMA (0, 1, 1) (1, 0, 0)* which can be expressed as

$$D(AV\_GR\_MONTH\_WAGE) = 0 + [AR(4) = 1.010423, MA(1) = 0.544233, BACKCAST = 2001Q2, ESTSMPL = "2001Q2\ 2011Q3"]$$

All model parameters are the 5% significance level statistically significant, the correspondence between model and data expressed by $R^2 = 0.946$ is sufficient and the corrected version by the degrees of freedom $\overline{R^2} = 0{,}945$ is also high. Using Breusch and Godfrey autocorrelation test, there was not rejected the null hypothesis about absence of autocorrelation at the 5% significance level. The ARCH test did not reject the null hypothesis testing homoskedasticity at the 5% significance level, so it is possible to claim that the model is estimated correctly as a whole.

By this model is possible to calculate the forecast for the 4th quarter of 2011, because this value will be officially published by the CZSO after the presentation of this article and next

one it is possible to calculate the forecast for subsequent 4 quarters of 2012. Even for this conditional forecast remain valid the same assumptions as in the model estimating the average number of persons employed. The published seasonally adjusted time series of average gross monthly wage per actual person by CZSO from 1Q2000 to 3Q2011 is displayed in Fig. 4 in red, the estimated forecast for 4Q2011 to 4Q2012 is displayed in blue. Better preview of trend will be further illustrated by graphic output with using seasonally adjusted series by X12 – ARIMA method.

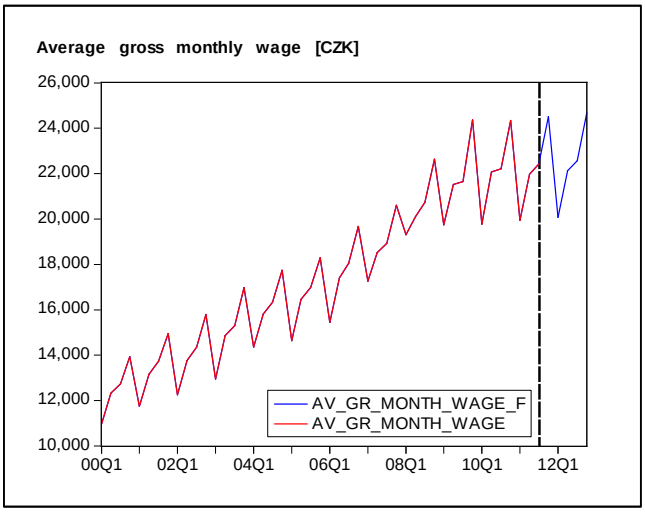


Fig. 4 – Forecast of average gross monthly wage per actual person, not seasonally adjusted. Source: CZSO - data, own construction – forecast.

In a subsequent section, we can prepare a vector autoregression model. The Granger hypothesis about absence of causality was rejected at the 5% significant level (see Tab. 7) and therefore we can say, that there is a causal relationship between the time series of average monthly gross wages per actual person and gross domestic product at constant prices.

Tab. 7 - Granger causality. Source: EViews, own construction.

VAR Granger Causality/Block Exogeneity Wald Tests			
Dependent variable: GDP_SA			
Excluded	*Chi-sq*	*df*	*Prob.*
AV_GR_MONTH_WAGE_SA	5.277062	1	0.0216
All	5.277062	1	0.0216
Dependent variable: AV_GR_MONTH_WAGE_SA			
Excluded	*Chi-sq*	*df*	*Prob.*
GDP_SA	5.692481	1	0.0170
All	5.692481	1	0.0170

There was constructed VAR(1) model and the estimated parameters are shown in Tab. 8. Using Breusch Godfrey autocorrelation test, there was not rejected the null hypothesis of absence autocorrelation at the 5% significance level and using White heteroscedasticity test there was not rejected the null hypothesis testing homoskedasticity at 5% significance level.

Tab. 8 - The estimated parameters of VAR(1) model. Source: EViews, own construction.

Vector Autoregression Estimates		
Standard errors in ( ) & t-statistics in [ ]		
	GDP_SA	*AV_GR_MONTH_WAGE_SA*
GDP_SA(-1)	1.051847	0.002323
	(0.01943)	(0.00097)
	[54.1435]	[2.38589]
AV_GR_MONTH_WAGE_SA(-1)	-1.994657	0.908492
	(0.86830)	(0.04351)
	[-2.29719]	[20.8784]
R-squared	0.963516	0.975491
Adj. R-squared	0.953369	0.965161
Sum sq. resids	2.89E+09	7247890.
S.E. equation	8098.930	405.8630
F-statistic	6742.104	2988.615
Log likelihood	-478.2252	-340.5255
Akaike AIC	20.87936	14.89241
Schwarz SC	20.95886	14.97192
Mean dependent	789165.7	17572.67
S.D. dependent	99456.15	3331.823

The VAR(1) model can be expressed as

$$GDP\_SA_t = 1.051847\ GDP\_SA_{t-1} - 1.994657\ AV\_GR\_MONTH\_WAGE\_SA_{t-1}$$

$$AV\_GR\_MONTH\_WAGE\_SA_t = 0.002323\ GDP\_SA_{t-1} + 0.908492\ AV\_GR\_MONTH\_WAGE\_SA_{t-1}$$

Seasonally adjusted time series of average gross wage per actual person is shown in Fig. 5 and Fig. 6 in red. In blue there are shown the forecasts for 4Q2011 – 4Q2012. The estimated values by VAR(1) model are shown in the left graph, the estimated values by SARIMA (1, 0, 0) (1, 0, 0)c model are shown in the right graph.

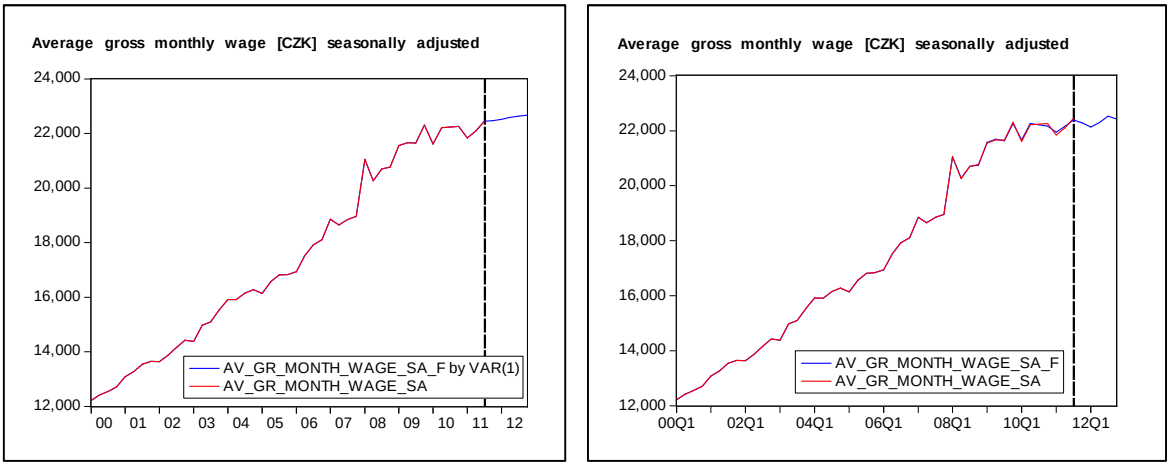


Fig. 5 (left), Fig. 6 (right) – Forecast of average gross monthly wage per actual person, (left – by VAR model, right – by SARIMA model), seasonally adjusted. Source: CZSO – data, own construction – forecast.

Both models provide approximately comparable estimates. It is possible that the wages of workers in construction in 2012 will slightly increase. This increase of "average" will be probably caused by **increase wages of particular groups of workers**, so the increase will not

be everywhere and uniform. The estimated values for each quarter are shown in Tab. 9 for the VAR model and in Tab. 10 for the SARIMA model.

Tab. 9 (left), Tab. 10 (right) – The estimated values of average gross monthly wage per actual person. Source: own construction.

OBS.	AV_GR_MONTH_WAGE_SA_F by VAR(1)
4Q2011	22 476.59
1Q2012	22 518.27
2Q2012	22 584.25
3Q2012	22 640.91
4Q2012	22 674.78

OBS.	AV_GR_MONTH_WAGE_SA_F
4Q2011	22 281.68
1Q2012	22 140.09
2Q2012	22 295.84
3Q2012	22 527.72
4Q2012	22 419.54

4 CONCLUSION

Due to high demand for timeliness of point estimates of important macroeconomic statistics, SARIMA and VAR models appear to be an effective tool. Simple estimates are calculated quickly and cheaply, and the resulting values are normally only slightly different. The reliability of these estimates is lower and lower when more and more unstable economy we consider. Czech economy belongs rather to more dynamic economies, but because the world still recovers from the economic slowdown that took place in 2009 and continued in 2010, it can not be expected any greater acceleration in the Czech economy for 2012. The forecasts, which are constructed by state offices and other analytical companies whose research budgets are expensive, provides more sophisticated results, however, these forecasts may be just as different from real values, if the sophisticated assumptions of these offices and companies are violated. The question is, how wide is the border between simple and sophisticated models and whether it would be better to more rely on speed and simplicity to the future, than waste resources on sophisticated analysis, which effect could disappear just as in case of simple analysis. The inserted resources of various sizes, however, in both cases may never be recovered.

References

1. BOX, G.E.P., JENKINS, G.: *„Time series analysis: Forecasting and control“*, San Francisco, Holden-Day, 1970.
2. BOX, G.E.P., PIERCE, D.A.: „Distribution of the Autocorrelations in Autoregressive Moving Average Time Series Models“, *Journal of the American Statistical Association*, 65, 1970, p. 1509–1526.
3. ENGLE, R.F.: *„ARCH: selected readings“*, Oxford University Press, 1995.
4. GRANGER, C.W.J.: „Investigating causal relations by econometric models and cross-spectral methods“, *Econometrica* 37 (3), 1969, p. 424–438.
5. HAMILTON, J.: *„Time Series Analysis“*, Princeton University Press, 1994.
6. HYLLEBERG, S., ENGLE, R.F., GRANGER, C.W.J., YOO, B.S.: „Seasonal Integration and Cointegration“, *Journal of Econometrics*, 44, 1990, p. 215-238.

Contact information

Ing. Ondřej Šimpach
University of Economics Prague, Faculty of Informatics and statistics
Winston Churchill sq. 4, 130 67 Prague 3
+420 224 095 273, +420 737 66 54 61
ondrej.simpach@vse.cz