PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1162 — 2007
Applications of Mathematics and Statistics in Economics

Jakub Fischer

University of Economics, Prague, Czech Republic

USING LEADING, COINCIDENT AND LAGGED SERIES AT PRELIMINARY ESTIMATES OF GDP: EXPERIENCES IN THE CZECH REPUBLIC

1. Introduction

Official estimates of quarterly changes of the gross domestic product (GDP) are at a disposal on 70th day after the reference quarter, so some earlier estimates are demanded. These preliminary estimates could also use the balances of the business cycle surveys (ratio of positive and negative expectations). In this paper we would like to present the model of preliminary estimates of GDP quarterly changes for the next quarter. Our research work follows the previous results of our colleagues. J. Jílek and M. Vojta [4;6] have constructed two alternative models for user's flash estimates of GDP changes using short-term production statistics. J. Jílek and K. Pikula [3] presented their experiences with using selected business cycle balances as explanatory variables in the model of GDP predictions. The fourth starting point for our modelling is the article of J. Jílek, M. Vojta and I. Pecáková [5]. It contains three parts – the first is focused on the predictions of sales (or production) based on business cycle surveys, the two other parts on deep analysis of micro-data from business surveys. They state that data from business tendency surveys are not very useful for predictions of production indicators.

In our work we would like to design a model for preliminary estimates of GDP quarterly changes, considering different types of series (leading, coincident and lagged series, represented by confidence indicators from business and consumers tendency surveys, natural indicators of electric energy consumption in kWh and rail transport of goods in tonne-kilometres, and finally by 2 lagged series of a dependent variable with different number of lags) as explanatory variables. Our pre-

liminary estimate, which should be at a disposal on the 30th day after the end of the reference period, is neither the flash estimate (flash estimates are based on the concept of national accounts and use the relations in the system of national accounts)[1] nor the leading indicator (leading indicators do not use any information related to the reference period). The model is designed in an aggregated form (it estimates only the aggregate, not individual components of resources or expenditures of GDP). The last-year-paper presented at 7th AMSE Conference in České Budějovice is developed and improved.

2. Methodology

As a common method for all the segments of the research we have selected the regression analysis. All the computations were provided on the basis of quarterly series. Monthly data were aggregated to quarters by the simple sum and then they were computed year-to-year changes of these quarterly aggregated series as well as of the series of quarterly GDP.

As a dependent variable we consider quarterly year-to-year changes (in percents) of the gross value added at basic constant prices at a price level of 1995 (hereinafter: H). As explanatory variables we consider quarterly year-to-year changes (in percent) of following series:

H1 – series H with a lag of 1 quarter,
H2 – series H with a lag of 2 quarters,
E – amount of electric energy consumption,
D – amount of rail transport of goods,
S1 – Confidence Indicator for Industry,
S2 – Confidence Indicator for Construction,
S3 – Confidence Indicator for Trade,
S4 – Business Confidence Indicator,
S5 – Consumer Confidence Indicator,
S6 – Composite Confidence Indicator.

The aim of regression analysis used is to select relevant explanatory variables and compute estimates of regression parameters relating to these variables. We used series from 1st quarter of 1999 till 3rd quarter of 2003. As an assessment of the models we use forecasts of the last value of a dependent variable (H), while the series used in the model have been shortened. It is possible to compare forecasts from shortened series with actual, but during the computation unknown, values of last quarter. As a criterion for an assessment of accuracy of the model there was used a difference of a model forecast from the official estimate published on the 70th day after the end of the reference period (i.e. from the last value of series H).

[1] *Handbook of Quarterly National Accounts* (1999), Chapter 16.

When we had computed the model, we took all the variables to the model and then we used forward and backward, respectively, selection methods of a stepwise regression. The aim of using mentioned selection methods was to reduce the number of explanatory variables in a model. After forward and backward selection we controlled the individual *t*-test of zero-values of regression parameter for each included variable. In the following steps we reduced the number of considered variables before starting forward and backward selection procedures; due to the correlation between some explanatory variables, using this approach we could find more than two combinations of explanatory variables (one combination from forward selection, one combination using backward selection). It is necessary to note that the aim of our work is not to find combinations of variables with a maximal coefficient of determination (when reducing of number of variables is not relevant), but to find combinations with the best prediction ability. This ability is assessed by forecasts from shortened series for the 3rd quarter of 2003.

Analogically as for the 3rd quarter, we firstly searched suitable variables for predictions of GDP growth for the other quarters of year 2003. We always considered data which were at our disposal in the time of constructions of shortened predictions (about 1 month after the end of the reference period). For each quarter we selected promising combinations of explanatory variables (using forward/backward selections), while the differences from later official estimates have been used as a criterion of "hopefulness".

Next these promising combinations were verified from the viewpoint of their prediction ability for the other quarters. Combinations were considered irrespective of an application of forward/backward selection methods. The aim of this plan was to verify, if the combinations suitable for predictions in given quarter could be also suitable for predictions in other quarters. It is not sufficient to consider an intersection of the models from individual quarters, due to two reasons: firstly, combinations for individual quarters were selected as samples, and some combinations could be omitted. The second reason was that the selecting model could exclude (or not include) a variable to the model for given quarter, which is necessary for another quarters. When we explore these combinations, which are feasible for all the quarters, eventual inclusion of nonessential variable does not lead to unacceptable error. While the combinations of variables are the same for all quarters, the vectors of regression parameters are computed separately for each quarter (recent data and relationships could be considered in this approach).

We also constructed the predictive models from series started at 2000, using a methodology mentioned above. Then we compared results (selected variables, coefficients of determination, biases of predictions) from 3-years series with previous results from 4-years series (the details of the results are in the paper [1]).

In the following phase, we verified a predictive ability of the combinations of explanatory variables on series started at 1999 for individual quarters of 2002 (pre-

viously we had found that 3-years series were sufficient for making predictions), so we obtained for every model shortened predictions for 7 quarters (in 2002, there were no official data in time due to the floods, which afflicted the building of the Czech Statistical Office in August).

Finally, we tried to design models with the lagged series of confidence indicator (of 1 quarter) and models using series from business and consumer surveys only.

3. Results

In Tables 1 and 2 there is a comparison of 10 best models based on a criterion of the average difference between preliminary estimate and official estimate (in percentage points). In Table 1 the models are sorted by mean absolute error, in Table 2 by maximal error.

In Table 3 there are estimates of parameters for the best model in each of 7 quarters, which contains variables D (rail transport), H1 (lagged dependent variable) and S3 (confidence indicator for trade); p-value for individual t-tests for each parameter is also stated.

We can see the comparison of predictions of preliminary and official estimates on the Fig. 1. The model uses strictly data which are at a disposal at a time of making preliminary estimate, so the series of GDP is not comparable in time. Official estimate for the 2nd quarter of 2002 were not released.

Table 1. Errors of estimates, prediction for 7 quarters of years 2002 and 2003, series started at 1999, sorted by mean absolute error in % points

Combination of explanatory variables	Minimal absolute error	Maximal absolute error	Mean error	Mean absolute error
D, H1, S3	0.030	0.663	–0.064	0.342
D, E, H2, S1, S6	0.124	0.677	–0.072	0.350
D, E, H2, S2	0.085	0.725	0.006	0.423
D, E, H1, H2, S1, S6	0.219	0.823	–0.141	0.467
H1, S2	0.078	1.171	–0.264	0.470
D, E, H1, H2, S5	0.034	1.071	–0.099	0.482
D, E, H1, H2, S6	0.242	0.988	0.229	0.482
D, E, H2	0.155	1.037	0.444	0.488
D, H1, S1, S4	0.047	1.421	–0.257	0.503
D, S1, S6	0.029	1.900	–0.436	0.536

Table 2. Errors of estimates, prediction for 7 quarters of years 2002 and 2003, series started at 1999, sorted by maximal absolute error in % points

Combination of explanatory variables	Minimal absolute error	Maximal absolute error	Mean error	Mean absolute error
D, H1, S3	0.030	0.663	–0.064	0.342
D, E, H2, S1, S6	0.124	0.677	–0.072	0.350
D, E, H2, S2	0.085	0.725	0.006	0.423
D, E, H1, H2, S1, S6	0.219	0.823	–0.141	0.467
D, E, H1, H2, S6	0.242	0.988	0.229	0.482
D, E, H2	0.155	1.037	0.444	0.488
D, E, H1, H2, S5	0.034	1.071	–0.099	0.482
H1, S2	0.078	1.171	–0.264	0.470
D, H1, S1, S5, S6	0.061	1.249	0.004	0.654
D, H1, S1, S4	0.047	1.421	–0.257	0.503

Table 3. Estimates of parameters of the best model

Period	β_0	β_1	p-value (β_1)	β_2	p-value (β_2)	β_3	p-value (β_3)	DW
2002:Q1	1.26	0.04	*	0.48	***	0.09	*	2.33
2002:Q3	1.35	0.04	*	0.45	***	0.09	*	2.44
2002:Q4	1.50	0.06	**	0.38	***	0.07	*	2.49
2003:Q1	1.72	0.07	**	0.45	***	–0.01	*	2.39
2003:Q2	1.45	0.06	**	0.38	***	0.08	**	2.38
2003:Q3	1.47	0.06	**	0.39	***	0.07	**	2.42
2003:Q4	1.57	0.07	***	0.39	***	0.06	**	2.47

Note: * mean p-value > 0.05; ** mean p-value is in interval (0.01; 0.05); *** means p-value < 0.01.

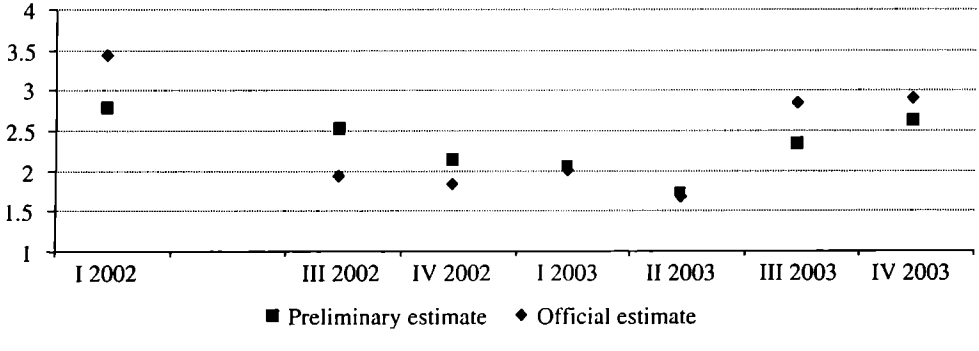

Fig. 1. Preliminary and official estimates of GDP quarterly year-to-year changes, %

In Tables 4 and 5 there is a comparison of models which alternatively use or do not use the shift of the series from business cycle surveys for 20 best models. Coefficient of determination as well as mean error are compared. The comparison is made for 2nd quarter of 2003 and 3rd quarter of 2003.

Table 4. Comparisons of preliminary estimates for the 2nd quarter of 2003 using shift of series from business cycle surveys of 1 quarter forward

Combination	Without shift		With shift		Differences		Result	
	Coef. of det.	Error	Coef. of det.	Error	Coef. of det.	Error	Coef. of det.	Error
D, S1, S4	79.31	–0.382	47.11	0.434	–32.20	0.052	–	–
S5	62.74	0.226	62.74	0.102	0.00	–0.124	–	+
D, E, H1, H2, S1, S6	77.68	–0.371	77.68	0.202	0.00	–0.169	–	+
D, E, H1, H2, S5	84.93	–0.008	83.23	–0.048	–1.70	0.040	–	–
D, E, H2	80.97	0.223	80.97	0.223	0.00	0.000	0	0
D, S1, S3	75.04	–0.159	34.91	1.297	–40.13	1.138	–	–
D, S2, S3, S4	83.02	–0.215	70.22	0.706	–12.80	0.491	–	–
D, E, H2, S2	84.72	–0.016	82.78	0.123	–1.94	0.107	–	–
H1, S2	71.64	–0.004	61.82	0.141	–9.82	0.137	–	–
D, H1, S1, S4	79.33	–0.397	51.16	0.768	–28.17	0.371	–	–
D, H1, S1, S5, S6	81.47	–0.331	89.30	0.773	7.83	0.442	+	–
D, H1, S3	65.61	0.101	59.78	1.220	–5.83	1.119	–	–
D, E, H2, S1, S6	86.14	–0.142	83.90	0.170	–2.24	0.028	–	–
D, S1, S5	72.70	0.328	72.71	0.170	0.01	–0.158	+	+
D, S1, S2	75.58	0.022	63.72	0.381	–11.86	0.359	–	–
D, S5, S6	70.21	0.418	74.42	0.225	4.21	–0.193	+	+
D, E, H1, H2, S6	81.62	0.132	82.91	0.004	1.29	–0.128	+	+
D, S1, S2, S3	81.93	–0.297	81.93	0.698	0.00	0.401	0	–
D, S1, S6	78.93	–0.152	60.72	0.208	–18.21	0.056	–	–

It is obvious from Tables 4 and 5 that shift of the series from business and consumers surveys does not lead to better models, measured both by the value of coefficient of determination and the error of estimate. It could be explained by the behaviour of respondents. They are affected by the current situation when they make their prediction, the second reason could be the better ability of firms to assess the current situation in comparison with making the prediction for following months.

The detailed table contains results of the models based on the business and consumer tendency surveys only are in the article [5].

Table 5. Comparisons of preliminary estimates for the 3rd quarter of 2003 using shift of series from business cycle surveys of 1 quarter forward

Combination	Without shift		With shift		Differences		Result	
	Coef. of det.	Error	Coef. of det.	Error	Coef. of det.	Error	Coef. of det.	Error
D, S1, S4	79.54	–0.039	52.28	–0.922	–27.26	0.883	–	–
S5	66.52	–0.823	55.30	–0.984	–11.22	0.161	–	–
D, E, H1, H2, S1, S6	88.08	0.252	85.46	0.712	–2.62	0.460	–	–
D, E, H1, H2, S5	86.06	0.106	84.54	0.069	–1.52	–0.037	–	+
D, E, H2	82.85	0.544	82.85	0.544	0.00	0.000	0	0
D, S1, S3	77.10	0.544	31.28	0.423	–45.82	–0.121	–	+
D, S2, S3, S4	83.82	–0.064	71.37	–0.345	–12.45	0.281	–	–
D, E, H2, S2	85.89	–0.021	84.48	0.017	–1.41	–0.004	–	+
H1, S2	74.28	–0.906	65.98	–1.300	–8.3	0.394	–	–
D, H1, S1, S4	79.59	–0.062	54.15	–0.609	–25.44	0.547	–	–
D, H1, S1, S5, S6	82.17	0.004	88.29	1.463	6.12	1.459	+	–
D, H1, S3	69.21	–0.427	54.17	0.410	–15.04	–0.017	–	+
D, E, H2, S1, S6	86.79	0.167	85.44	0.684	–1.35	0.517	–	–
D, S1, S5	75.39	–0.019	75.60	–0.360	0.21	0.341	+	–
D, S1, S2	78.00	–0.392	67.29	–0.800	–10.71	0.408	–	–
D, S5, S6	72.81	–0.202	77.11	–0.154	4.3	–0.048	+	+
D, E, H1, H2, S6	83.40	0.377	84.35	0.236	0.95	–0.141	+	+
D, S1, S2, S3	82.35	–0.104	71.38	–0.366	–10.97	0.262	–	–
D, S1, S6	80.39	–0.093	64.90	–0.830	–15.49	0.737	–	–

4. Conclusion and plans for further work

In the recent work we found an acceptable model for predictions of gross value added at 30th day after the end of the period. The model contains leading, coincident and lagged series. We might consider our partial results as satisfactory, due to biases of official estimates which we analysed and presented in the 6th AMSE Conference in 2003 [2].

There are some general findings on modelling of GDP predictions in the Czech Republic. It is more suitable to estimate gross value added as an aggregate, not by its components. Assessment of the models is better using the difference between preliminary and official estimate instead of the value of the coefficient of determination. Series of length more than 5 years are not suitable, 4-years-series models and 3-years-series models are very similar. Shift of the series from BCTS does not lead to better results. Confidence indicators are useless when used separately.

For further work, we plan to improve the model from the gross value added (in basic prices) to the gross domestic product in purchasing prices and estimate the GDP changes for 1-2 next quarters by estimating of the explanatory variables.

References

[1] Fischer J., *Gross Domestic Product: To the Optimal Length of Time Series Used in Predictive Models*, Statistics: Investment in the Future. Praha 6-7.09.2004.

[2] Fischer J., "Using Obstacles of GDP Estimates in the Czech Republic. Banská Bystrica 04.09.2003 - 05.09.2003". [in:] *Applications of Mathematics and Statistics in Economy. Acta Oeconomica No. 16.* Univerzita Mateja Bela, Ekonomická fakulta, Banská Bystrica 2004, pp. 43–47.

[3] Fischer J. et al., *Řady konjunkturních sald jako vysvětlující proměnné řad relativních změn HDP ČR. Průběžná zpráva o plnění dílčího úkolu v roce 2002*, [výzkumná zpráva], VŠE FIS, Praha 2002, 69 pp.

[4] Jílek J., "Zdokonalené uživatelské signální odhady čtvrtletních změn hrubého domácího produktu České republiky", *Politická ekonomie* 2003, (5), 676–694.

[5] Jílek J., Pecáková I., Vojta M., "Konjunkturní průzkumy ČSÚ a predikce", Statistika 2005, 5 (Praha).

[6] Jílek J., Vojta M., "User Flash Estimates of Short-term Changes in Gross Domestic Product of the Czech Republic", *Statistics in Transition* 2001, 5 (2), 281–300.

STOSOWANIE SZEREGÓW WIODĄCYCH, WSPÓŁWYSTĘPUJĄCYCH I OPÓŹNIONYCH WE WSTĘPNYCH OSZACOWANIACH PKB: DOŚWIADCZENIA Z REPUBLIKI CZESKIEJ

Streszczenie

W artykule omówiono wstępne oszacowania PKB. Stosując ten model, możliwe jest oszacowanie wzrostu PKB 30 dni po kwartale odniesienia (w porównaniu z 70 dniami koniecznymi do oficjalnych oszacowań). Model dotyczy wiodących, współwystępujących i opóźnionych szeregów. Jest weryfikowany przez prognozy z krótkich szeregów. Dyskutowana jest optymalna długość szeregu czasowego stosowanego w modelu, a także możliwość korzystania wyłącznie z danych z sondażów biznesowych i konsumenckich. Jakość prognozy jest porównana z jakością oficjalnych oszacowań PKB.

Słowa kluczowe: wstępne oszacowania PKB, sondaże tendencji biznesowych i konsumenckich, spójność szeregów jakościowych i ilościowych.