STATISTICS IN TRANSITION new series, September 2017
Vol. 18, No. 3, pp. 481–500, DOI: 10.21307/stattrans-2016-082

OPTION FOR PREDICTING THE CZECH REPUBLIC'S FOREIGN TRADE TIME SERIES AS COMPONENTS IN GROSS DOMESTIC PRODUCT

Luboš Marek[1], Stanislava Hronová[2], Richard Hindls[3]

ABSTRACT

This paper analyses the time series observed for the foreign trade of the Czech Republic (CR) and predictions in such series with the aid of the SARIMA and transfer-function models. Our goal is to find models suitable for describing the time series of the exports and imports of goods and services from/to the CR and to subsequently use these models for predictions in quarterly estimates of the gross domestic product (GDP) component resources and utilization. As a result we get suitable models with a time lag, and predictions in the time series of the CR exports and imports several months ahead.

Key words: transfer-function models, SARIMA models, quarterly estimates of the Gross Domestic Product (GDP), imports and exports of goods and services, exchange rate.

1. Introduction

Imports and exports of goods and services are ranked among the most important economic indices. The data of imports and exports describe the economic relationships from the viewpoint of goods and services circulation among residents and non-residents (in this sense, we consider the so-called national concept to foreign trade, based on the goods circulation as the change in ownership. Another possible approach is that of the so-called cross-border one, i.e. the principle of the goods passing the state borders); they express the extent to which the economy is open (regarding GDP). In this respect, they not only play an indispensable role for assessing economic performance, but are also significant components in the GDP estimate by the expense method.

[1] Department of Statistics and Probability, University of Economics, Prague, Czech Republic. E-mail: marek@vse.cz.
[2] Department of Economic Statistics, University of Economics, Prague, Czech Republic. E-mail: hronova@vse.cz.
[3] Department of Statistics and Probability, University of Economics, Prague, Czech Republic. E-mail: hindls@vse.cz.

Short-term (quarterly) estimates of GDP are based on two fundamental methods related to the two fundamental data sources: the production method is based on the estimates of the gross value added created in individual industries of the national economy, i.e. the GDP resources; and the expense method stems from the estimated components of GDP, i.e. final consumption, gross capital formation, and net imports. By making the estimates more accurate and better balanced, we get to the sole GDP value informing us about the economy evolution in the past quarter.

As data collection and processing become faster, and the economic evolution more turbulent during the year, the demands put forth by the users of statistical data are naturally growing with respect to the speed and quality of the short term estimates. These two parameters, speed and quality, go in opposite directions from the viewpoint of governmental statistics; if the user requires *quality* (i.e. as accurate as possible) statistical information, it must be derived from rather extensive surveys. Naturally, such surveys need time for preparation, collection and processing of data and, logically, such procedures are summarily used for annual data, or annual national accounts. If the user requires data *as quickly as possible* upon the completion of the period under consideration, each statistical office must necessarily take into account limited data sources and prefer modelling to direct findings. As s result we get quick estimates of the economy evolution during the year, but the user must allow for the estimates to be subsequently made more accurate. The quarterly estimates must be a kind of a trade-off between the speed and quality. If the GDP components are concerned, they must provide, as quickly as possible, reliable information on economic circulation during the year. The criterion of this reliability should be according to the ability to provide the source information for the initial estimates of the GDP annual values that would ensure the smallest possible deviations from the annual values based on extensive annual surveys. Hence, the quality of estimating the GDP components (imports and exports of goods and services, final consumption, and gross capital formation) plays the decisive role in this respect. Demands have been ever-growing for how quickly the initial information on the GDP evolution should be provided. Currently, it is expected that the initial estimates on the GDP growth rate should be available within 30 days from the end of the respective quarter in all EU countries; a more accurate GDP estimate together with the structure of resources and utilization within 60 days from that date; and the complete sector accounts within 90 days from that date.

The goal of the present paper is to provide a new, original methodological support to those quick model estimates – here they are focused on the estimates of imports and exports of goods and services as significant components of the expense method for the GDP estimate. The proposed original methodology will enable us not only to estimate the values for the imports and exports of goods and services, but also to calculate their estimates quicker, as source data for estimating the quarterly GDP. This model has been verified on the data for the imports and exports of goods and services to/from the Czech Republic found in the database of the Czech Statistical Office (the data is stated in tsd. CZK) and the data of the exchange rate evolution taken from the database of the Czech National Bank. The analysis was carried out in the SCA software and eViews software.

2. Formulation of Problem

The issues connected with the short-term estimates for macroeconomic aggregates are inseparably connected with the effort to provide users of statistical data with reliable information about the evolution of the national economy as quickly as possible. Should such information be consistent, it must necessarily be viewed within a wider context of the system of macroeconomic statistical data, that is, the national accounts. In this sense, the short-term estimates are only relevant for a limited number of macroeconomic aggregates, namely, those entering the relationships in the GDP creation and utilization, based on the production and expense methods for estimating GDP (we are not going to consider the context of completing the quarterly national accounts in this paper; cf. e.g. Eurostat (2013a), Eurostat (2013b) or Marini (2016) for more details).

The GDP quarterly estimate, therefore, is based, on the one hand, on estimated gross value added, created in individual industries of the national economy (the estimates for resources, i.e. the gross value added in individual industries in the Czech Republic, are considered in the paper by Marek *et al.* (2016)) and, on the other hand, on the estimated expense components (imports and exports of goods and services, final consumption and gross capital formation) with subsequent balancing so that the sole value of the quarterly GDP is achieved.

For modelled (indirect) estimates of macroeconomic aggregates, a number of approaches can be utilized, stemming from the methods for time series analysis or regression analysis taking into account the relationships between annual and quarterly values. These methods are:

- without the quarterly or monthly data in the form of a reference index,
- using a reference index.

The estimating methods not utilizing a reference index enable us to get preliminary estimates of quarterly values, exclusively using formal mathematical procedures and criteria, providing smoothed quarterly estimates fulfilling a constraint that the sum of quarterly values over all four quarters equals the respective annual value. In other words, the annual value is disaggregated into quarters on the basis of purely formal criteria, without any knowledge about the evolution of the chosen index (or other indices) during the year. The best-known methods of this type include BFL (cf. e.g. Boot *et al.* (1967) or Wei and Stram (1990), Al-Osh (1989)). The usual models of time series (ARIMA) can also be viewed as members of this group. Such methods should only be used if a suitable reference index cannot be established, and for less important values. A natural utilization of them would also be a correction of an estimate obtained in the first step of desegregation with the aid of a reference index (for example, one variant of the BFL method is based on minimizing the sum of squares of second differences, which criterion can also be used for corrections of an estimate after the first step).

The methods with the aid of a reference index make use of external quarterly, or even monthly, information about a related index (or several related indices). The main feature of these procedures is the use of a quarterly or monthly established index that is factually tied with the value of an annual aggregate, to facilitate the distribution of the annual value into quarterly ones. Mathematically, a formal (regression) model is used for the relationship between the annual value of the aggregate to be estimated and a quarterly (usually average) value of the reference index. This model makes it possible to get an initial estimate of the aggregate; in the second step this estimate is corrected so that the sum of all four quarterly values equals the annual value of the aggregate (in particular, the INSEE method is used, created by Bournay and Laroque (1979); more information about this method can be found in, e.g. Nasse (1973) or Dureau (1991); approaches making use of a dynamic variable can also be classified into this group, e.g. Moauro and Savio (2005) or Mitchell *et al.* (2005)).

In parallel with this traditional method of disaggregation with subsequent correction, methods of disaggregation without subsequent corrections have been developed and utilized; that is, methods which establish already at the first step such estimates for the aggregate's quarterly values that their sum complies with its annual value (this group of models contains, for example, those described by Chow and Lin (1971), and Kozák *et al.* (2000)).

Indisputably, the core aspect of the estimation quality in this group of methods is finding a suitable reference index. In the range of all indices coming into consideration, i.e. fulfilling the above-stated conditions, we have to seek for the one that best corresponds to the short-term evolution of the aggregate in question. This suitability should be observed on a prolonged time horizon and separately for each aggregate whose values are to be estimated with the aid of the disaggregation method. This stage requires thorough analytical work, which must not be either neglected or underestimated. When applying indirect methods, we need not only to create a formal statistical model but also to build up an entire system of short-term statistical data.

Another category of short-term estimating methods is focused on predicting quarterly values of aggregates (regardless of their relationships to the annual values) with the aid of methods used in time series analysis. Having in mind the nature of the problem, we can use classical decomposition, linear dynamic models or spectral analysis (more details of these methods can be found in, e.g. Green (2008), Pankratz (1991), Wei (2006), Anderson (1976), and Granger and Newbold (1986), Proietti (2011), Bikker (2013)). This paper offers one option for the utilization of the available short-term survey results in deriving estimates of the quarterly values from the underlying model. As a result of this approach, based on the analysis of time series, we obtain a stable and factually relevant model for estimating the quarterly values of imports and exports of goods and services as components in the expense method for estimating GDP while using a suitable reference index. On a long-term basis, this model can be used for estimating quarterly values of other aggregates concerning the formation and use of GDP provided that the reference index is chosen appropriately.

This model is based on the utilization of the reference index and a time lag, due to which unknown values of imports and exports of goods and services can be estimated even without knowledge of the value of the reference index in the current (i.e. currently estimated) or future periods.

In this article, we focused precisely on a quick estimate of imports and exports of goods and services. This is especially due to the significance of foreign exchange for the Czech economy. Imports of goods and services in the Czech Republic is currently around 83% of GDP and exports of goods and services, about 77% of GDP (it is recalled that exports of goods and services are in FOB prices and imports of goods and services at CIF prices). The development of the values of these indicators is very closely related to the phases of the economic cycle because the basis of exports is mainly the products of the manufacturing industry and the basis of the import of raw material. For this reason, a timely and reliable estimate of the value of imports and exports of goods and services can be used to make a faster estimate of the quarterly GDP.

Imports and exports of goods and services are monitored in the Czech Republic in the so-called national concept. This is in line with the national accounts methodology, i.e. with the concept of other macroeconomic indicators. The national concept of foreign trade statistics follows up the actual trade in goods carried out between Czech and foreign entities, i.e. trade, where there is a change of ownership between residents and non-residents. Thus, the national concept of foreign trade reflects the export and import performance of the Czech economy better than the so-called cross-border concept. Conversely, the cross-border conception of foreign trade only reflects the physical movement of goods across the Czech Republic, irrespective of whether there is trade between Czech and foreign entities. These data, which only refer to the physical movement of goods from and into the territory of the Czech Republic, are surveyed for the purpose of international comparison of the movement of goods and services. However, they are not suitable indicators for monitoring the development of the economy in relation to GDP growth.

3. Methodology and results

In each of the above-mentioned groups classifying the time series analysis, there are many other approaches, and choosing from among them is governed by the character of the underlying data. For the time series we encounter here, we have selected a combination of models from the areas of stochastic methods and linear dynamic models to describe not only the behaviour of each time series separately, but also how values of one time series depend on those of another, and to reflect this relationship in the model as well. As a prerequisite to the utilization of such a model in efficient short-term predictions, the time lag must be incorporated. When selecting suitable models, we viewed ARIMA and SARIMA ones (cf. Anderson (1976), Box *et al.* (1994), Granger and Newbold (1986), Wei (2006)), as well as transfer-function models (cf. Pankratz (1991), SCA Statistical System (1991)).

The sources of our data were the monthly series of imports and exports of goods and services to/from the Czech Republic (in tsd. CZK and current prices; source: www.czso.cz), and the time series of monthly average values of the CZK/EUR exchange rates (source: www.cnb.cz**)**. These series were available from January 1999 to September 2016 for imports and exports (213 observations), and from January 1999 to October 2016 for exchange rates (214 observations). The complete analysis was carried out in the SCA software. The values of the imports and exports of goods and services are published monthly by the Czech Statistical Office and, due to the character of this data (relationships to the quarterly and annual national accounts) they are reviewed several times. On the other hand, the values of the CZK/EUR exchange rates (monthly averages) are published by the Czech National Bank immediately upon the end of the respective month and are not reviewed any more.

We consider the stochastic models of time series in their general form:

$$\phi_p(B)\Theta_P(B^L)(1-B)^d(1-B)^D Y_t = \theta_q(B)\Theta_Q(B^L)\varepsilon_t \tag{1}$$

where Y_t is the output series, ε_t is the random variable (white noise), B is the shift operator ($BY_t = Y_{t-1}$), L is the length of season, p is the order of AR process, q is the order of MA process, P is the order of seasonal AR process, Q is the order of seasonal MA process, d is the order of differencing, D is the order of seasonal differencing, ϕ_p is the autoregressive operator of order p, θ_q is the moving average operator of order q, Θ_P is the seasonal autoregressive operator of order P and Θ_Q is the seasonal moving average operator of order Q (cf. e.g. Box, Jenkins, Reinsel (1994)).

For model identification, the values of the autocorrelation and partial autocorrelation functions (ACF, PACF), and also the extended and inverse autocorrelation functions (EACF, IACF) were mainly used. The output is very extensive; hence only the most important aspects are explicitly mentioned. All the time series under analysis were non-stationary and had to be transformed to achieve stationarity (mostly by current and seasonal differentiating). The stationarity was tested by several approaches – the unit root, homoscedasticity, and Dickey-Fuller tests.

The values of the cross-correlation function (CCF) were calculated to prove the linear dependence between the analysed (already transformed, i.e., stationary) time series. These values confirmed the linear dependence between the transformed series of imports and exports of goods and services to/from the

Czech Republic and the transformed series of exchange rates. Afterwards, the general transfer-function model was applied:

$$Y_t = c + \nu_0 X_t + \nu_1 X_{t-1} + \nu_2 X_{t-2} + ... + \nu_K X_{t-K} + \frac{1}{(1-\phi_1(B))(1-\Phi_1(B^L))}\varepsilon_t \qquad (2)$$

where Y_t is the output series of exports, or imports (after the relevant transformations); X_t is the input series of the CZK/EUR exchange rates (again after the relevant transformations); and the last term is the perturbation series, denoted by N_t in the literature. The LTF method – cf. Pankratz (1991) and SCA Statistical System (1991) – was used for the parameter estimates. The resulting model was used for the predictions. The quality of the predictions was evaluated with the aid of the Theil coefficient of inequality.

$$TIE = \frac{\sum_{i=1}^{T}(\widehat{Y}_t - Y_t)^2}{\sum_{i=1}^{T}(Y_{t-1} - Y_t)^2} \qquad (3)$$

When predicting, we first shortened the analysed time series and created the so-called *dormant predictions*, i.e. predictions for the periods in which we had already known the actual values of the time series. This approach enabled us to compare the predictions with the actual values and thus to assess the model quality in the most objective way. Only afterwards we predicted for several periods ahead. We cannot aim at a too ambitious horizon for the predictions because the economic conditions under which the time series evolves are quickly changing in time. Moreover, predictions for longer horizons are not necessary with respect to the nature of the problem in question.

3.1. CR Exports

The time series of the CR exports (denoted by *Exports* in our analyses) is seasonal and non-stationary. We applied current and seasonal differentiation to the series to achieve stationarity. The tests carried out confirmed our approach.

First of all, a SARIMA model suitable for this series was established. After a thorough analysis and study of ACF, PACF, EACF, IACF, and unit root, homoscedasticity, and Dickey-Fuller tests, we identified our model as (cf. the SCA output):

$$(1-0.3856B^3 + 0.177B^{10})Y_t = (1-0.6119B)(1-0.5628B^{12})\varepsilon_t$$

where $Y_t = (1-B)(1-B^{12})\,Export_t$, ε_t is the white noise, and B is the classical backward-shift operator ($B^k Y_t = Y_{t-k}$). This model was successful at all stages of verification and was proven as fully adequate; this fact is also indicated by the value of the index of determination, which amounts to 0.983.

Table 1. Model parameter estimates for the output series (SCA software output)

```
SUMMARY FOR UNIVARIATE TIME SERIES MODEL Exports
-----------------------------------------------------------------------
VARIABLE   TYPE OF    ORIGINAL     DIFFERENCING
          VARIABLE   OR CENTERED
EXPORTN    RANDOM     ORIGINAL     (1-B  ) (1-B  )
-----------------------------------------------------------------------
PARAMETER   VARIABLE  NUM./  FACTOR  ORDER   CONS-    VALUE      STD     T
   LABEL      NAME    DENOM.                 TRAINT              ERROR  VALUE
  1     TH    EXPORTN    MA      1       1     NONE     .6119     .0635   9.64
  2   TH12    EXPORTN    MA      2      12     NONE     .5628     .0686   8.21
  3   PHI3    EXPORTN    AR      1       3     NONE     .3856     .0723   5.33
  4  PHI10    EXPORTN    AR      1      10     NONE    -.1770     .0693  -2.55

EFFECTIVE NUMBER OF OBSERVATIONS . .          185
R-SQUARE . . . . . . . . . . . . . .         .983
RESIDUAL STANDARD ERROR. . . . . .   .100513E+08
```

Subsequently, we determined the model for the CZK/EUR exchange rate series (denoted by *EUR* below). The corresponding SARIMA model is:

$$(1-B)X_t = (1-0.2186B)\varepsilon_t$$

where $X_t = EUR_t$, with the index of determination at 0.991. It is clear from the model that current differentiating was used to achieve stationarity.

Table 2. Model parameter estimates for the input series (SCA software output)

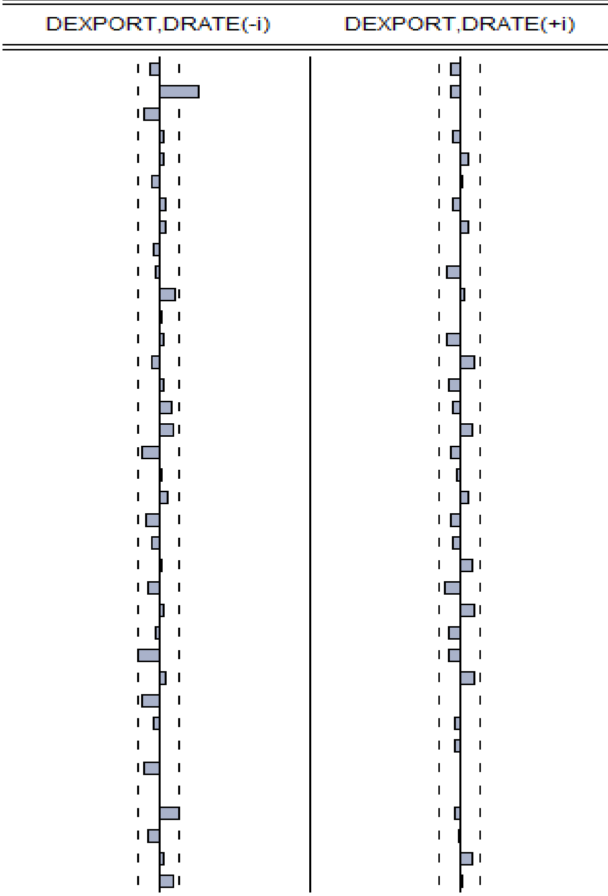
```
SUMMARY FOR UNIVARIATE TIME SERIES MODEL --   EUR
-----------------------------------------------------------------------
VARIABLE   TYPE OF    ORIGINAL     DIFFERENCING
          VARIABLE   OR CENTERED
                                        1
  EURN     RANDOM     ORIGINAL     (1-B  )
-----------------------------------------------------------------------
PARAMETER   VARIABLE  NUM./  FACTOR  ORDER   CONS-     VALUE     STD     T
   LABEL      NAME    DENOM.                 TRAINT              ERROR  VALUE
   1   PHI      EURN     AR       1      1     NONE     .2186     .0633   3.45

EFFECTIVE NUMBER OF OBSERVATIONS . .          206
R-SQUARE . . . . . . . . . . . . . .         .991
RESIDUAL STANDARD ERROR. . . . . .   .364515E+00
```

Another output from the SCA software shows the values of the cross-correlation function between (now already stationary) time series $(1-B)(1-B^{12})Y_t$ and $(1-B)X_t$. The Cross-Correlation Function (CCF) values indicate not only the intensity of mutual linear dependence between the differentiated series, but also the direction of that dependence.

Both the values and the curve imply what the significant value of CCF is (95% confidence interval) at time *t*-1. We identified significant linear dependence between the transformed time series of exports at time *t* and the transformed time series of the CZK/EUR exchange rates at time *t*-1.

Correlations are asymptotically consistent approximations



i	lag	lead
0	-0.0654	-0.0654
1	0.2995	-0.0598
2	-0.1062	0.0029
3	0.0392	-0.0497
4	0.0326	0.0646
5	-0.0424	0.0219
6	0.0572	-0.0566
7	0.0437	0.0583
8	-0.0389	0.0151
9	-0.0170	-0.0912
10	0.1223	0.0299
11	0.0214	0.0134
12	0.0261	-0.0842
13	-0.0544	0.1061
14	0.0263	-0.0730
15	0.0873	-0.0410
16	0.0944	0.0850
17	-0.1116	-0.0586
18	0.0228	-0.0176
19	0.0647	0.0635
20	-0.0888	-0.0598
21	-0.0429	-0.0430
22	0.0177	0.0879
23	-0.0669	-0.0995
24	0.0340	0.0978
25	-0.0241	-0.0667
26	-0.1385	-0.0766
27	0.0554	0.0935
28	-0.1156	0.0036
29	-0.0259	-0.0293
30	0.0132	-0.0263
31	-0.0998	0.0087
32	0.0008	0.0132
33	0.1371	-0.0361
34	-0.0779	-0.0103
35	0.0317	0.0826
36	0.0915	0.0179

Figure 1. CCF evolution (95% confidence interval), eViews output

Next we identified the transfer-function model. The LTF method – cf. Pankratz (1991) – was used for this identification. The value of the v_1 weight was the only one that was significantly different from zero; it means that the values of the output series of exports depend on those of the input series of the CZK/EUR exchange rates with a time lag equal to one. This fact had already been indicated by the CCF values. The remaining weights were identified as insignificant by our testing. After a thorough analysis we had thus established a suitable model and estimated its parameters. The resulting transfer-function model hence is:

$$(1-0.4387B^3+0.1394B^{10})Y_t =$$
$$=\text{-}5{,}706{,}000+7{,}253{,}000 * X_{t\text{-}1}+(1-0.6316B)(1-0.5541B^{12})\varepsilon_t$$

where $Y_t=(1-B)(1-B^{12})\,Export_t$ and $X_t=(1-B)*EUR_t$.

Table 3. Parameter estimates for the TFM model (SCA software output)

```
SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- EXPORT1
-----------------------------------------------------------------------
VARIABLE   TYPE OF    ORIGINAL     DIFFERENCING
          VARIABLE   OR CENTERED
                                        1      12
EXPORTN    RANDOM     ORIGINAL     (1-B  )  (1-B  )
                                        1      12
  EURN     RANDOM     ORIGINAL     (1-B  )  (1-B  )
-----------------------------------------------------------------------
PARAMETER  VARIABLE  NUM./  FACTOR  ORDER   CONS-     VALUE      STD      T
   LABEL      NAME    DENOM.                TRAINT               ERROR  VALUE
  1    V0      EURN    NUM.     1      0     NONE -.5706E+07  .165E+07  -3.45
  2    V1      EURN    NUM.     1      1     NONE  .7253E+07  .166E+07   4.37
  3   PHI    EXPORTN    MA      1      1     NONE      .6316     .0634   9.97
  4  PHI12   EXPORTN    MA      2     12     NONE      .5541     .0698   7.94
  5   PHI3   EXPORTN    AR      1      3     NONE      .4387     .0705   6.22
  6  PHI10   EXPORTN    AR      1     10     NONE     -.1394     .0659  -2.12

EFFECTIVE NUMBER OF OBSERVATIONS . .        185
R-SQUARE . . . . . . . . . . . . . . .      .984
RESIDUAL STANDARD ERROR. . .. . .    .956100E+07
```

This model was successfully verified (with the aid of the above-mentioned procedures and methods) and established as fully adequate. Its quality is, among other things, confirmed by the value of the index of determination, amounting to 0.984. The CCF values between the residuals of the SARIMA and transfer-function models did not significantly differ from zero, which can be seen in the output cited below. In other words, the model we established complied with one of the most important verification criteria.

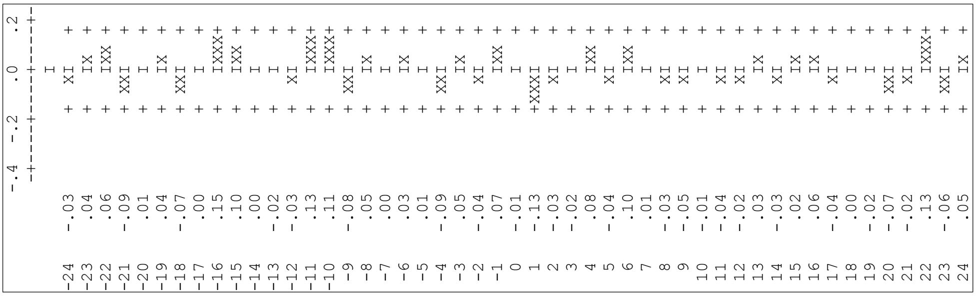


Figure 2. Evolution of CCF residuals (95% confidence interval)

Therefore, we can say that the value of the time series of imports (after the current and seasonal differentiating) at time *t* depends on the past values of the series itself (with time lag values of 3 and 10), the values of the time series of the CZK/EUR exchange rates (after current differentiating) at time *t*–1, and the past values of the random component (with a seasonal parameter).

Let us now have a look at the predictions because of which both of the above-described estimates were predominantly derived. As of the time of writing the present paper, the values of the time series of exports are known up to August 2016, but those of the exchange rates up to October 2016. Therefore, we were able to make use of the time lag, and input into the transfer-function model not predictions (which would be usual) but the actually observed values. Of course, we expected improvement of the predictions from this step. The situation is illustrated in the following Figure:

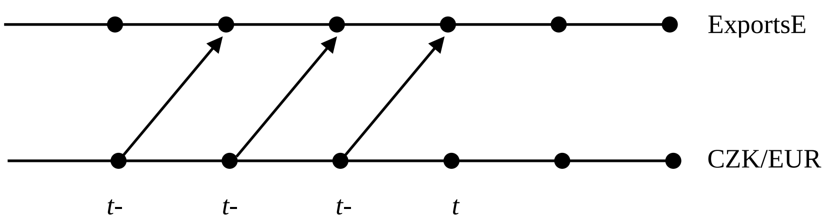


Figure 3. Linear relationship between the exports and CZK/EUR exchange rates

3.2. Predictions for Exports

Table 4 shows the five-month-ahead predictions of the shortened time series of exports as derived with the aid of the SARIMA and the transfer-function (TFM) models. Thanks to our shortening the time series by five values, we can compare the predictions with the actual values.

Table 4. Exports – predictions versus actual values

	Predictions		Actual values
Month	SARIMA	TFM	
May	300,450,000	307,651,000	328,285,676
June	319,340,000	329,600,000	349,080,472
July	287,780,000	287,840,000	277,464,937
August	286,780,000	296,590,000	309,977,136
September	316,340,000	326,160,000	347,200,140

Table 5 sums up the standard deviation values of the predictions. The comparison between the predicted and actual values is satisfactory (cf. Table 6). Somewhat more accurate predictions and smaller values of the standard deviation are in favour of the TFM model.

Table 5. Exports – standard deviations of predictions

	Standard deviations	
Month	SARIMA	TFM
May	9,274,300	9,221,400
June	9,978,900	9,948,600
July	1,063,700	1,063,100
August	1,268,200	1,260,600
September	1,355,400	1,354,800

Table 6. Exports – comparison of predictions versus actual values

	Difference (prediction minus actual value)		Ratio (prediction/actual value)	
Month	SARIMA	TFM	SARIMA	TFM
May	-27,835,676	-20,634,676	0.885	0.937
June	-29,740,472	-19,480,472	0.858	0.944
July	10,315,063	10,375,063	1.037	1.037

August	-23,197,136	-13,387,136	0.861	0.957
September	-30,860,140	-21,040,140	0.882	0.939

Table 7. Theil coefficient of inequality

SARIMA	TFM
0.3985	0.1879

The Theil coefficient of inequality clearly indicates that TFM is better.

After comparing the models with the actual values, we decided for the transfer-function one. Making use of the full available length of the time series, we predict five months ahead. The results are shown in Table 8. Of course, the actual values are unknown to the authors at the time of writing this paper; hence the accuracy can only be measured after five additional months.

Table 8. Exports – predictions

Month	TFM
October	336,240,000
November	360,020,000
December	299,350,000
January	314,050,000
February	332,310,000

3.3 CR Imports

Let us now analyse the time series of imports. The procedure is identical with the one we used for exports. Again, we created the SARIMA model, this time for the time series of imports, and the TFM one – imports depending on the CZK/EUR exchange rates. We made the predictions and compared them.

Here, we only state the particular models and predictions, without detailed reasoning and software output (except for the resulting TFM model).

3.4 Predictions for Imports

The following formula describes the suitable SARIMA model for the time series of imports:

$$(1-0.3368B^3-0.2126B^5+0.2968B^{10})Y_t=(1-0.6029B)(1-0.6622B^{12})\varepsilon_t\,,$$

where $Y_t = (1-B)(1-B^{12})\,Import_t$, and ε_t is the white noise. This model successfully passed the complete verification stage. The index of determination equals 0.978 – hence we would hardly be able to find a better model.

The SARIMA model for the series of the CZK/EUR exchange rates was already identified when constructing the model for exports. We can directly continue to the construction of the TFM model. According to the SCA output stated below, the TFM formula is:

$$(1-0.3331B^3-0.2036B^5+0.3227B^{10})Y_t = 1{,}070{,}000 * X_{t\text{-}1} + (1-0.6289B)(1-0.7207B^{12})\varepsilon_t$$

where $Y_t = (1-B)(1-B^{12}) * Import_t$ and $X_t = (1-B) * EUR_t$

This model also passed successfully all stages of the verification procedure and was found fully adequate, with the value of the index of determination amounting to 0.977. Table 10 shows the prediction values and Table 11 their standard deviation values.

Table 9. Parameter estimates for the TFM model (SCA software output)

```
SUMMARY FOR UNIVARIATE TIME SERIES MODEL -- IMPORTS1
-----------------------------------------------------------------------
VARIABLE   TYPE OF    ORIGINAL     DIFFERENCING
           VARIABLE   OR CENTERED
                                        1       12
 IMPORT    RANDOM     ORIGINAL     (1-B   )  (1-B   )
                                        1
  EURN     RANDOM     ORIGINAL     (1-B   )
-----------------------------------------------------------------------
PARAMETER  VARIABLE  NUM./  FACTOR ORDER   CONS-     VALUE       STD     T
   LABEL       NAME    DENOM.              TRAINT                 ERROR  VALUE

  1    V1     EURN    NUM.     1      1     NONE  .1070E+07  .834E+06   1.28
  2    TH     IMPORT   MA      1      1     NONE      .6289     .0620  10.15
  3   TH12    IMPORT   MA      2     12     NONE      .7207     .0528  13.65
  4   PHI3    IMPORT   AR      1      3     NONE      .3331     .0703   4.74
  5   PHI5    IMPORT   AR      1      5     NONE      .2036     .0677   3.01
  6  PHI10    IMPORT   AR      1     10     NONE     -.3227     .0691  -4.67

EFFECTIVE NUMBER OF OBSERVATIONS . .         190
R-SQUARE . . . . . . . . . . . . . .        .977
RESIDUAL STANDARD ERROR. . . . .. .   .962642E+07
```

Table 10. Imports – predictions versus actual values

	Predictions		Actual values
Month	SARIMA	TFM	
May	290,450,000	289,300,000	286,759,551
June	299,340,000	299,600,000	302,126,184
July	287,780,000	272,840,000	250,955,952
August	266,780,000	269,590,000	275,437,737
September	306,340,000	305,160,000	298,809,680

At first sight we can see that the SARIMA and TFM predictions do not differ from each other to a great extent. However, the TFM predictions are closer to the actual values – this can also be seen in Table 12. In Table 11, we can see the standard deviations of the predictions, which are again smaller for TFM.

Table 11. Imports – standard deviations of predictions

	SARIMA	TFM
May	9,274,300	9,021,400
June	9,978,900	9,548,600
July	1,062,700	913,100
August	1,262,800	1,005,600
September	1,395,400	1,054,800

Table 12. Imports **–** comparison of predictions versus actual values

	Difference (prediction minus actual value)		Ratio (prediction/actual value)	
Month	SARIMA	TFM	SARIMA	TFM
May	3,690,449	2,540,449	1.013	1.009
June	-2,786,184	-2,526,184	0.991	0.992
July	36,824,048	21,884,048	1.147	1.087

August	-8,657,737	-5,847,737	0.969	0.979
September	7,530,320	6,350,320	1.025	1.021

Table 13. Theil coefficient of inequality

SARIMA	TFM
0.3772	0.1415

The Theil coefficient of inequality clearly indicates that TFM is better.

For the imports, the transfer-function model was also established as better. We made the five-month-ahead predictions within an approach identical to that for the exports. The results are shown in Table 14.

Table 14. Imports – predictions

Month	TFM
October	962,640,000
November	314,880,000
December	270,950,000
January	278,410,000
February	288,920,000

The above-presented model clearly shows that the theory of stochastic models for time series was used (whether SARIMA or the theoretically more demanding TFM models); this theory by itself is more complex than, for example, the often-used decomposition of time series. Both models provide good results. However, the transfer-function model (TFM) is better with respect to accuracy of predictions, values of the standard deviation of the estimates, as well as to the Theil coefficient of inequality. For these reasons, the authors prefer the TFM model despite the fact that this model is more complex. Our analysis has clearly established that using simpler models is out of the question due to the nature of the data. Of course, the analysis itself is very demanding and laborious for the same reason.

Both SARIMA and TFM are relatively complex models. Those who deal with such models know very well that it would certainly be possible (having in mind the duality between auto-regression and moving averages) to identify more models with a different (or similar) structure of independent variables, and such models could be used for predicting alternatively. However, such models would hardly be simpler and there is a question whether better predictions would be achieved using them. The authors deal with such models and predictions based on them on a long-term basis. On the basis of empiric experience it can be confirmed that the given models are robust, i.e. relatively stable in time regarding the individual variables (components) in the model. It means that the parameter estimates are changing in time, but the models as such remain stable regarding the structure of the variables.

When the present paper is being published, other data have been published for the foreign-trade time series (the data for several months have been modified and/or added; and later data for approximately the last two years have been changed within the framework of regular reviews of the quarterly and annual national accounts). But the data for the exchange rates will remain unchanged. A question hence arises whether or not our predictions become useless. The answer is, of course, that they do not: the above-described predictions are valid for the given data and model and are important at the time of being calculated, because they can be used for subsequent analyses, source information for decision-making, and future considerations. The validity of predictions and reviews of the original data represent a far more general problem; this problem is valid for most data published by each and every statistical office. What is most important in this context is the model robustness. If the model is robust, it can be used even when the data changes, recalculating just the values of the parameters and predictions.

4. Conclusions

The goal of the present paper is to establish models for the time series of exports and imports of goods and services from/to the Czech Republic suitable for the construction of short-term predictions. Our analysis has proved mutual linear dependence between monthly time series (exports and imports) of the foreign trade of the Czech Republic, expressed in million CZK in current prices, and the time series of the CZK/EUR exchange rates (monthly averages).

Within the framework of our analysis, we created suitable SARIMA models for all the time series concerned. We have also derived transfer-function models for the series of exports and imports, in which the input time series were those of the CZK/EUR exchange rates. When predicting in a TFM model, predictions of the input series are usually used, on the basis of which predictions of the output series are calculated. Thanks to the linear dependence (with a time shift) between the series we have proved within the analysis, as well as the different times of publishing the time series values (the exchange rate values are published several months earlier than those of the foreign trade), we can utilize the actual values of the input series (i.e. the exchange rates) instead of their predictions in the TFM model. This approach, logically, leads to better quality of predictions for the output series (exports or imports for our case). All these facts can be utilized in estimating the evolution of the CR foreign trade on a short time horizon of two to three months. Nevertheless, it should be noted that (and this comment is even more important at the time of a crisis) predictions of economic time series are purposeful only a few periods ahead because the external influences on their evolution are quickly changing, thus in principle disabling the creation of good-quality long-term predictions. Moreover, we should realize that the time series values of the CR foreign trade published on the website of the Czech Statistical Office are just initial estimates for the most recent periods. Such initial estimates are subsequently reviewed and made more accurate (related to the reviews of the quarterly and annual national accounts). On the other hand, the models described in the present paper enable us, at a relatively high level of quality, to provide estimates of closely watched economic time series that substantially affect quarterly estimates of GDP.

The authors feel that the importance of such an analysis is characterized not so much by publishing the actual values of the predictions but rather by the methods, procedures and models used – those can be instructive for predicting in time series of the foreign trade of the Czech Republic, as important components of quick estimates of GDP.

Acknowledgements

This paper was written with the support of the Czech Science Foundation project No. P402/12/G097 "DYME – Dynamic Models in Economics" and with the support of the Institutional Support to Long-Term Conceptual Development of Research Organization, the Faculty of Informatics and Statistics of the University of Economics, Prague.

REFERENCES

AL-OSH, M., (1989). A Dynamic Linear Approach for Disaggregating Time Series Data, Journal of Forecasting, Vol. 8, pp. 85–96.

ANDERSON, O. D., (1976). Time series analysis and forecasting – Box-Jenkins approach, London, Butterworth.

BIKKER, R., DAALMANS, J., MUSHKUDIANI, N., (2013). Benchmarking large accounting frameworks: a generalized multivariate model, Economic Systems Research, Vol. 25, pp. 390–408.

BOOT, J. C. G., FEIBES, W., LISMAN, J. H. C., (1967). Further Method of Derivation of Quarterly Figures from Annual Data, Applied Statistics, Vol. 16, pp. 65–75.

BOURNAY, J., LAROQUE, G., (1979). Réflexions sur la méthode d´élaboration des comptes trimestriels, Annales de l'INSEE, Vol. 11, pp. 3–18.

BOX, G. E. P., JENKINS, G. M., REINSEL, G. C., (1994). Time series analysis, forecasting and control, Third edition. Prentice-Hall, Inc. Englewood Cliffs, New Jersey.

DENTON, F. T., (1991). Adjustment of monthly and quarterly series to annual tools: An approach based on quadratic minimization, Journal of American Statistical Association, Vol. 66, pp. 99–102.

DUREAU, G., (1991). Les comptes nationaux trimestriels, Paris: INSEE – Méthodes.

FORNI, M., HALLIN, M., LIPPI, M., REICHLIN, L., (2005). The generalized factor model: one-sided estimation and forecasting. Journal of the American Statistical Association, Vol. 100, pp. 830–840.

GRANGER, C. W. J., NEWBOLD, P., (1986). Forecasting Economic Time Series, Academic Press, New York.

GREENE, W. H., (2012). Econometric Analysis, Pearson Education, Prentice Hall.

EUROSTAT, (2013a). European System of Accounts (ESA 2010), Eurostat, Luxembourg.

EUROSTAT, (2013b). Handbook on quarterly national accounts, Eurostat, Luxembourg.

KOZÁK, J., HINDLS, R., HRONOVÁ, S., (2000). Some Remarks to the Methodology of the Allocation of Yearly Observations into Seasons, Statistics in Transition, Vol. 4, pp. 815–826.

MAREK, L., HRONOVÁ, S., HINDLS, R., (2016). Příspěvek k časnějším odhadům hodnot čtvrtletních národních účtů [Contribution to the earlier estimations of quarterly national accounts], Politická ekonomie, Vol. 64, pp. 633–650.

MARINI, M., (2016). Nowcasting Annual National Accounts with Quarterly Indicators: An Assessment of Widely Used Benchmarking Methods. IMF Working Paper 16/71.

MITCHELL, J., SMITH, R. J., WEALE, M. R., WRIGHT, S., SALAZAR, E. L., (2005). An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth. Economic Journal, Vol. 115, pp. 108–129.

MOAURO, F., SAVIO, G., (2005). Temporal Disaggregation Using Multivariate Structural Time Series Models, Econometrics Journal, Vol. 8, pp. 214–234.

NASSE, PH., (1973). Le système des comptes nationaux trimestrieles, Annales de l´INSEE, Vol. 5, pp. 119–161.

PANKRATZ, A., (1991). Forecasting with dynamic regression models, John Wiley & Sons Inc., New York.

PROIETTI, T., (2011). Multivariate temporal disaggregation with cross-sectional constraints, Journal of Applied Statistics, Vol. 38, pp. 1455–1466.

THE SCA STATISTICAL SYSTEM, (1991). Reference manual for general statistical analysis, Scientific Associates Corp. Oak Brook, Illinois, USA.

WEI, W. W. S., (2006). Time series analysis – univariate and multivariate methods, Pearson, Addison Wesley Publishing, New York.

WEI, W. W. S., STRAM, D. O., (1990). Disaggregation of Time Series Models, Journal of Royal Statistical Society, Vol. 52, pp. 453–467.