CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Bachelor thesis

2018 **Jana Učňová**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Jana Učňová

Predicting Czech Economic Activity Using Toll Data

Bachelor thesis

Prague 2018

Author: Jana Učňová

Supervisor: Ing. David Kocourek

Academic Year: 2017/2018

Bibliographic note

UČŇOVÁ, Jana. *Predicting Czech Economic Activity by Toll Data.* Prague 2018. 35 pp. Bachelor thesis (Bc.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis supervisor Ing. David Kocourek

Abstract

Many analysts coincide that transportation is closely linked to economic activity. However, data containing information about transportation have not been part of their research for a long time. Introduction of electronic toll collection systems in recent years led to a new source of data containing information about truck transport. This thesis aims to examine the ability of seasonally adjusted toll data to predict Czech economic activity. Economic activity is represented by four variables - real GDP, nominal GDP, industrial production index and the volume of foreign trade. Seven models - five dynamic models, ARIMA model, and regression with ARIMA error - are constructed for each dependent variable. These models are then compared using both Akaike and Bayesian information criterion and the most appropriate model for each dependent variable is selected. It was concluded that both real GDP and industrial production index can be predicted using toll data. Both the number of kilometers travelled, and the amount of toll collected seems to be good predictors of economic activity. Particularly, data containing information about toll collected might be more beneficial because the amount of toll collected in given quarter can even predict economic activity in the next quarter.

Keywords

nowcasting, economic activity, toll data, electronic toll collection, dynamic model, ARIMA, Akaike information criterion, Bayesian information criterion

Abstrakt

Mnoho analytiků se shoduje v tom, že doprava je úzce spojena s ekonomickou aktivitou. Nicméně, data obsahující informace o dopravě dlouhou dobu nebyla součástí jejich výzkumů. Zavedení elektronických systémů výběru mýtného vedlo v nedávné době k novému zdroji dat, která obsahují informace o kamionové dopravě. Jsou tedy užitečným zdrojem infromací, který může být použit k předpovědi ekonomické aktivity. Cílem této práce je zkoumání schopnosti sezónně očištěných dat z mýtných bran předpovídat ekonomickou aktivitu České republiky. Ekonomická aktivita je zastoupena čtyřmi proměnnými - reálným HDP, nominálním HDP, indexem průmyslové produkce a objemem zahraničního obchodu. Sedm modelů - pět dynamických modelů, ARIMA model a regrese s ARIMA chybou - jsou sestraveny pro každou závislou proměnnou. Tyto modely jsou poté srovnány pomocí Akaikeho a Bayesova informačních kritérií a pro každou závislou proměnnou je vybrán ten nejvhodnější model. Jak reálné HDP, tak i index průmyslové produkce lze předpovědět pomocí mýtných dat a počet projetých kilometrů i vybraného mýta se zdá být dobrým prediktorem ekonomické aktivity. Zejména, vybrané mýto může být přínosnější, jelikož lze pomocí něj předvídat ekonomickou aktivitu dokonce i v následujícím kvartálu.

Klíčová slova

nowcasting, ekonomická akticita, mýtná data, elektronické mýtné, dynamický model, ARIMA, Akaikeho informační kritérium, Bayesovo informační kritérium

Declaration of Authorship

I hereby proclaim that I wrote my bachelor thesis on my own under the leadership of my supervisor and that the references include all resources and literature I have used.

I grant a permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, 11 May 2018

Signature

Acknowledgment

I would like to express my gratitude especially to Ing. David Kocourek for his valuable advices and useful comments. Furthermore, I would like to thank to Ing. Petr Vaculčík from Road and Motorway Directorate of the Czech Republic for help with data collection. Finally, I would like to thank to Ing. Petr Učeň and Martin Frič MSc for their ideas, help and support.

Bachelor's Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author's name and surname: Jana Učňová

E-mail: jana.ucnova@seznam.cz

Phone: +420607669670

Supervisor's name: Ing. David Kocourek

Supervisor's email: kocourek.david@email.cz.

Proposed Topic:

Predicting Czech Economic Activity Using Toll Data

Preliminary scope of work:

Research question and motivation

The form of the toll system and its possible targets of use are actual topics in the Czech Republic. Highways and expressways began to be charged on January 1, 1995. The new toll system was introduced in 2007 and since that vehicles with a total weight greater than 12 tonnes have been obliged to pay an electronic toll for highways, expressways and selected sections of first-class roads. Since 2010 the obligation to pay electronic toll has been extended to vehicles with a total weight greater than 3.5 tonnes and buses have been added since 2011. The toll rates are based on emission class of given vehicle and its number of axles.

Introduction of electronic toll collection system led to a new source of data containing information about truck transport. This data can be therefore a useful source of information which might be used to predict economic activity. Authors like Askitas and Zimmermann (2013) and Döhrn (2013) used toll data from Germany, and they found the strong correlation between transportation and business cycles. Especially, month-on-month variation in kilometers travelled appeared as a good predictor of German industrial production index.

Electronic collection system provides a large amount of data which can be used. The aim of this thesis is to examine the ability of toll data to predict Czech economic activity.

In my thesis, I would like to focus on following research questions.

- Can the economic activity be predicted using toll data?
- Which of the variable containing information about truck transport can best predict economic activity?
- Are effects of toll data on economic activity positive or negative?

Contribution

Askitas and Zimmermann (2013) and Döhrn (2013) examined the effect of toll data on German industrial production index. To my knowledge, no authors focused on a similar analysis in the Czech Republic. Adding other dependent variables like real GDP, nominal GDP, and trade balance, some other significant effects might be found out. Thus, this thesis might serve as a framework for how to proceed in selecting the most appropriate model predicting economic activity, which can be then used in different states of Europe.

Methodology

I am going to be using time series data from 2007 to 2017 which shows the number of kilometers travelled, the number of vehicles travelled and the amount of toll collected among highways and roads. As a dependent variable, real GDP, nominal GDP, industrial production index and balance trade will be used. I am going to construct several models for each dependent variable, which will be later on compared using Akaike information criterion and the most appropriate model for each dependent variable will be selected.

Outline

1. Introduction
2. Literature review
   - 2.1. Indicators of economic activity
   - 2.2. Transportation and the economy
   - 2.3. Electronic toll collection
   - 2.4. Toll data as a tool of nowcasting
3. Methodology
   - 3.1. Dynamic models
   - 3.2. ARIMA models
   - 3.3. Regressions with ARIMA error
4. Results
5. Conclusion

List of academic literature:

Bibliography

ASKITAS, Nikolaos; ZIMMERMANN, Klaus F. Nowcasting business cycles using toll data. Journal of Forecasting, 2013.

DÖHRN, Roland. Transportation data as a tool for nowcasting economic activity–The German road pricing system as an example. 2013.

LAHIRI, Kajal; YAO, Vincent Wenxiong. Economic indicators for the US transportation sector. Transportation Research Part A: Policy and Practice, 2006.

YAO, Vincent W. The causal linkages between freight and economic fluctuations. International Journal of Transport Economics/Rivista internazionale di economia dei trasporti, 2005.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. Forecasting: principles and practice. OTexts, 2014.

Author **Supervisor**

Contents

Introduction

During the last decades, many analysts have been trying to find the new way how to nowcast economic activity. Nevertheless, indicators containing information about transportation have not been included in their analysis for a long time. One reason for that was a relatively long lag in data publication. However, the availability of such data has considerably improved in recent years due to the introduction of electronic toll collection systems which allow the data to be better accessible (Döhrn, 2013).

Efforts to use data created by toll collection systems have increased and, particularly in the United States, various studies, which analyze the connection between transportation and the economy, have been released (Lahiri et al., 2004, Yao, 2005). In Europe,most of the analysis focused on German economic activity. Authors like Askitas and Zimmermann (2013) and Döhrn (2013) found the strong correlation between toll data and economic activity. They observed the ability of toll data to nowcast German industrial production index by which GDP can be identified. However, especially in case of predicting industrial production, the advantage of the model using toll data against other models is not so large because the difference in release of toll data and industrial production index is relatively small (Döhrn, 2013).

Electronic toll collection system has been operating within the territory of the Czech Republic since 2007 (Černý, 2008). It produces a huge amount of data which can be used in many fields. The 10-years long time series containing information about toll collection can be applied similarly as in the case of Germany. Thus, this is the main motivation of this thesis because none of the authors focused on the effects of information derived from toll data on the economic activity in the Czech Republic.

Hence, the main objective of this thesis is to explore the effect of truck transportation on economic activity in the Czech Republic which is represented by real GDP, nominal GDP, industrial production index and balance of foreign trade. Adding another three dependent variables might contribute to previous studies because studies from Germany concentrated only on predicting industrial production index. According to Askitas and Zimmermann (2013) and Döhrn (2013), it is expected that the effect of the number of kilometers travelled will be the most significant compared to other two variables containing information about truck transport. Furthermore, it is supposed that all these three variables - the number of kilometers travelled, the number of vehicles travelled and the amount of toll collected - will have a positive effect on economic activity in the Czech Republic.

This thesis mainly follows an approach used in studies by Askitas and Zimmermann (2013) and Döhrn (2013); that is why different sets of regressions are constructed which are then compared with each other, and the preferable model for each dependent variable is selected. As in the study by Döhrn (2013), ARIMA errors are included to assess whether models including toll data have better ability to predict economic activity than ARIMA models. The main hypothesis are as follows:

- Variables containing information about truck transportation can predict economic activity represented by real GDP, nominal GDP, the volume of foreign trade and industrial production index.

- The effects of variables containing information about truck transportation on economic activity are positive.

- There is at least one significant effect of variables containing information about truck transportation in the preferable model.

- Dynamic models including variables derived from toll data are more accurate for predicting economic activity than ARIMA models.

- Information about truck transportation are relevant and should be included in regression with ARIMA error.

The thesis is organized as follows. Chapter 1 summarizes literature and theoretical background related to the thesis. Chapter 2 describes data used in the thesis and provides seasonal adjustment and testing stationarity. Chapter 3 specifies methodology used, Chapter 4 provides results of thesis research, and the last Chapter concludes.

Chapter 1

Literature review and theoretical background

This chapter reviews literature related to the topic of the thesis and provides a theoretical background used later on. Firstly, the chapter focuses on research that mentions a variety of indicators which are used to predict business cycles and economic activity. Especially, timely high-frequency electronic indicators showed up to be remarkably informative about the current state of economy (Carlsen and Storgaard, 2010). Secondly, the role of transportation and its relation to economy are provided. In the third part, the chapter points out a road infrastructure charging of heavy-goods vehicles in Europe and explains Eurovignette Directive of the European Parliament and the Council. Next, electronic toll system in the Czech Republic is briefly described. Finally, the interesting position of toll data and its possibilities of usage in different fields of interest are shown. The chapter continues to mention literature about the application of toll data as a tool of nowcasting economic activity which is considered to be a relatively new trend because of the introduction of electronic toll collection only several years ago.

1.1 Indicators of economic activity

The lag in availability of relevant data is considered to be the main issue in forecasting and monitoring the aggregate economy (Edey et al., 1991). In European Union, Eurostat regularly publishes flesh estimates of quarterly GDP 30 days after the quarter-end since the end of April 2016 (Eurostat, 2018a, Kokkinen and Wouters, 2018). Edey et al. (1991) reviewed the performance of two groups of indicators - major expenditure aggregates from the quarterly national accounts and partial indicators - that were used in analyzing economic activity in Australia using VAR systems. Camba-Mendez et al. (2001) constructed an automatic leading indicator model, which uses a dynamic factor model to summarize the information about a variety of variables, and they used data from France, Germany, Italy and United Kingdom. The results showed that this approach appears to be more accurate than VAR models because an automatic leading indicator model avoids arbitrary nature of both forecasting VAR models and methods of constructing traditional leading indicators.

A large number of authors focused on usage of stocks and bonds to predict real economic activity. Estrella and Hardouvelis (1991) found out that the change in real GNP can be predicted by the slope of the yield curve and they pointed out that steeper slope of yield curve implies more rapid real output growth in the future. They used quarterly data from United States from 1955 to 1988 and showed that yield spread, the difference between 10-year government bond rate and 3-month T-bill rate, can forecast marginal changes in GNP up to one year and a half in the future, while in case of prediction of cumulative changes in GNP it is even up to 4 years in the future. This work was followed by Rosenberg and Maurer (2008) who divided term spread into two components - expectation and term premium. They constructed probit models to forecast whether the recession will occur in 12 months. The sample period was from July 1961 to July 2006 and during this period six recessions were defined by National Bureau of Economic Research, i.e., a private, non-profit organization which focuses on analyzing effects of public policies and estimating models of economic behaviour and then widening its economic research findings among other professionals (The National Bureau of Economic Research, 2018). Their work showed that expectation component is leading indicator of oncoming recession. Aylward and Glen (2000) used data from 23 countries, from which 15 of them were developing, to check out the possibility of using stock market prices to predict economic growth of income, consumption, and investment. The analysis showed the evidence that upturn in stock prices is generally pursued by an increase in GDP, consumption, and investment.

In recent years, electronic transaction data, which can be determined on a daily basis, or even in real time, started to have an essential role in the prediction of economic activity. Thus, there was a need to try to find a range of more timely, but less complete, indicators to predict current conditions of the aggregate economy. Carlsen and Storgaard (2010) used electronic transaction data by Dankort, a debit card developed by Danish banks, to examine whether electronic payments can be used as an indicator for retail sales in Denmark. They showed that electronic transaction can help with nowcasting and they can forecast retail sales, thus also GDP, in the short-term. Gill et al. (2012) analyzed the ability to estimate current economic conditions by using wholesale and retail electronic transaction data and internet search data by Australian households and businesses. Their results showed that wholesale and retail payments contain appropriate information about economic indicators and it pointed out that both payment behavior and internet usage become more stable. Thus they can be used by official statisticians to measure real-time economic aggregates. Moreover, Aprigliano et al. (2017) noticed that payment data - cheques, direct transfer, credit transfer and payment cards - can play a significant role in short-term forecasting of economic activity.

1.2 Transportation and economy

Having both reliable forecasting tools and real-time data, which can be used to analyze the current phase of the business cycle, is considered to be valuable (Guzavicius et al., 2013). Gordon (1992) found new measures of multi-factor productivity growth in rail, air and truck transportation that were later on compared with official ones from National Income and Product Accounts, and Bureau of Economic Analysis data. Han and Fang (2000) performed four measures - transportation industry's GDP, transportation final demand, transportation-related GDP, and transportation-driven GDP - which are statistically commensurate with the GDP and together determine a significance of transportation in the economy of United States.

Nonetheless, Lahiri et al. (2004) pointed out the critical role of transport in business cycles using data from the United States. They tested synchronization of cycles of four composite coincident indicators - transportation service index (TSI), real payrolls of workers in the transportation sector, real personal consumption expenditures on transportation services, all employees of transportation sector - and reference cycle using concordance index and correlation of cycles among these indicators. The evidence of co-movement of the four coincident indicators with reference cycle and application of both dynamic factor models and the nonparametric NBER procedure led to the estimation of the composite coincident index (CCI) that is considered to be the root for determination of a current state of the aggregate economy. They also pointed on the need to see the difference between full-fledged recessions of the economy, generally bordered by the time of slow growth, on one hand and standalone growth slowdowns, which do not come to full-fledged recession, on the other.

Studying business and growth cycles of the transportation sector in the United States made an exciting conclusion. "Relative to the economy, business cycles in the transportation sector have an average lead of 6 months at peaks and an average lag of nearly two months at troughs. Thus, the recessions in this sector last longer by nearly eight months than those of the overall economy." (Lahiri et al., 2004). Yao (2005) followed the previous work and examined the linkages between freight transportation and economic activity. He found that freight transportation is closely associated with inventory management and industrial production by using stage-of-fabrication model.

1.3 Charging heavy-goods vehicles in Europe

Eurovignette Directive is directive on charging heavy-goods vehicles (above 3.5 tonnes), and it was originally introduced in 1999[1]. It gives a framework for standard rules on distance-related toll and time-based user charges (vignettes) for heavy-goods vehicles for using roads that are part of trans-European network. The framework targets on reducing the differences between systems of toll and vignettes in the Member States that ensures both fair and efficient road pricing and the regular functioning of the EU transport market. It can also contribute to reduce CO_2 emissions and other transport's external costs which can lead to reducing air pollution. Time-based vignettes does not reflect the actual road use; thus they are taken as insufficient to conduct the efficiency of transport. Therefore distance-based charging is considered to be far preferable in enhancing transport behaviour (Kenny, 2017).

The application of road pricing is not obligatory for the Member States of the EU; however, when Member State decides to apply toll or vignettes, it needs to follow specific rules to levy such charges. For instance, the toll must be charged in accordance with the distance travelled and the type of vehicle, while vignettes are levied according to the duration of the use of appropriate infrastructure and the emission class of the vehicle. Both toll and vignettes must be non-discriminatory[2]. Within Europe, many countries, like Germany, Austria and Switzerland, introduced road tolling schemes for trucks and these systems differ across the EU (McKinnon, 2006).

In the Czech Republic, electronic toll collection (ETC) was introduced in

[1] Directive 1999/62/EC later modified by Directive 2006/38/EC and Directive 2011/76/EU

[2] European Comission: MEMO/10/489 (2010), European Comission: Mobility and Transport (2018)

January 2007. All heavy vehicles with a total weight of 12 tons and more are obligated to pay an electronic toll for usage of highways, motorways and selected 1st class roads. This event followed after Czech Government's series of discussion which started shortly after the country joined the EU in 2004 (Černý, 2008). The main reasons of introduction of ETC were considered to be an acquisition of higher volume of funds to initial development of infrastructure, an increase in competitiveness between road and rail transport and a reduction of the number of trucks (Ministry of Transport, 2018). Similarly as in France, Italy, Spain or Portugal, the Czech Republic toll all heavy-goods vehicles by road segment using toll plazas. This type of HGV tolling system uses a dedicated short-range communication (DSRC) technology which is regarded to be less costly (Broaddus and Gertz, 2008). From the January 2010, the ETC was extended on vehicles with total weight above 3.5 tons. Furthermore, in September 2011, buses were also added to ETC. The toll rates vary across the number of axes of given vehicle, emission class to which it falls and the type of infrastructure. Furthermore, in 2010, the new toll rates were introduced which require all heavy-goods vehicles to pay higher toll on Friday from 15:00 to 21:00. This peak hours rate might reflect the seasonality of the data produced by toll collection systems (Ministry of Transport, 2018, MYTO.cz, 2018).

1.4 Toll data as the tool of nowcasting

Toll collection can be seen not only as a source of money for government officials; however, it is also the vast amount of data which can be easily used in many fields, especially they are valuable for economists as well. For instance, it was expected an increase in truck traffic due to the new Amazon warehouse at Prague Airport. Based on the data from the toll system, the government can easily find out how much transport flow in the Czech road network increases due to this warehouse. They can also map not only transport near this warehouse but also they can see where trucks come in and out from the territory of the Czech Republic. The same findings can be detected in case of the extension of Škoda Auto plant (Šitner, 2017). Ministry of Transport of the Czech Republic prepares the new system which can give the most accurate information to drivers about traffic on domestic roads and motorways, and at the best case, it can lead them elsewhere to prevent traffic congestion (Sůra, 2018). Moreover, it can use them to monitor where truck drivers spend their mandatory stops and due to provide a more detailed capacity evaluation of motorway rest areas, and then possibly design new ones Šitner (2017).

Askitas and Zimmermann (2011) introduced the Toll Index, the new type of indicator for German business cycle forecasting. Four variants of Toll Index based on monthly number of kilometers travelled, the number of trips made, the number of inbound and outbound trucks to nowcast German Industrial Production Index (GPI). They followed the study by Bodo et al. (1991) who used electricity consumption data to forecast the level of industrial production in Italy. Askitas and Zimmermann (2013) found out that they can, based on Toll Index, nowcast GPI, which is an indicator of GDP, around one month ahead. It makes Toll Index a potential early indicator for business cycle forecast. Döhrn (2013) pointed at the power of toll data to predict German economic activity. He showed that model containing the number of kilometers travelled by vehicles is considered to be more accurate for predicting industrial production than ARIMA model. He concluded that the seasonally adjusted number of kilometers travelled appears to be a good predictor of production dynamics and toll data contribute further insights to business cycle analysis.

Chapter 2

Data

This chapter describes data used in the thesis and provides detection of their trends. Then the chapter continues with the identification of seasonality, and seasonal adjustment of toll data is presented. Moreover, stationarity is described and tested. All data used for this research are time series with the period from January 2007 to December 2017.

2.1 Toll data and dependent variables

Toll data used in the thesis are collected by Road and Motorway Directorate of the Czech Republic and are available at VÝROČENKY.cz (2018) - an online database which provides statistical data about transport and economy of transport. Overview about toll collection are procurable since 2007. They contain not only the number of kilometers travelled by all heavy-goods vehicles and the amount of toll collected within the territory of the Czech Republic, but also distribution by country of registration, number of axles, emission class and, furthermore also type of infrastructure.

In this thesis, three variables containing information about truck transport are used[1]. Firstly, the variable *vehicles* gives the number of heavy-goods vehicles which passed more than 0 kilometers in a given month within the territory of the Czech Republic. It indicates the sum of all vehicles independently of the nationality, number of axles and emission class. The variable

[1]Summary statistics about toll data are shown in Table C2 in Appendix C.

vehicles varies between 110,218 and 264,353 vehicles per month with its minimum on February 2007, maximum on October 2017 and mean that account for 197,678. Secondly, the variable *km* demonstrates the total number of kilometers travelled by all heavy-goods vehicles in a given month within the Czech Republic. That means the sum of kilometers travelled on highways, motorways and 1st class roads. The range of variable *km* lies between 95,857,517 and 246,520,999 kilometers with its minimum on December 2007, maximum on October 2017 and mean of 166,073,642 kilometers. Finally, the variable *toll* provides the total toll collected from all heavy-goods vehicles in a given month within the Czech Republic. In contrast with variable *km*, variable *toll* also displays the number of axles of given vehicle and its emission class because toll rates differ across these two indicators; thus this variable demonstrates a variety of traffic. It ranges between 389,284,675 and 961,631,079 CZK with its minimum on January 2008, maximum on October 2017 and mean that comprises of 666,033,022 CZK.

Figure 2.1 illustrates plots of variables described above - *vehicles*, *km* and *toll*. Graphically, it can be seen that in all three cases, the upward trend is present. The identification of a monotonic trend is considered to have high importance (Rutkowska, 2015); thus Cox-Stuart test is applied to verify this intuition. The alternative hypothesis in Cox-Stuart test is that the time series is characterized by a monotonic - upward or downward - trend. Using Cox-Stuart test[2], the p-value equals indeed to zero for all three variables - *vehicles*, *km* and *toll*. Thus, the null hypothesis is rejected and it is confirmed that the monotonic trend genuinely exists in all three cases.

As the first two dependent variables, quarterly real and nominal GDP of the Czech Republic are used and they are expressed in millions of CZK. Czech Statistical Office (2018) uses three methods to compute GDP. GDP used in this thesis is applied by expenditure method as the sum of the final use of products and services by resident units and the balance of exports and imports of goods and services.

The volume of foreign trade, which is defined as the external balance

[2]Results from Cox-Stuart test can be seen in Table D1 in Appendix D.

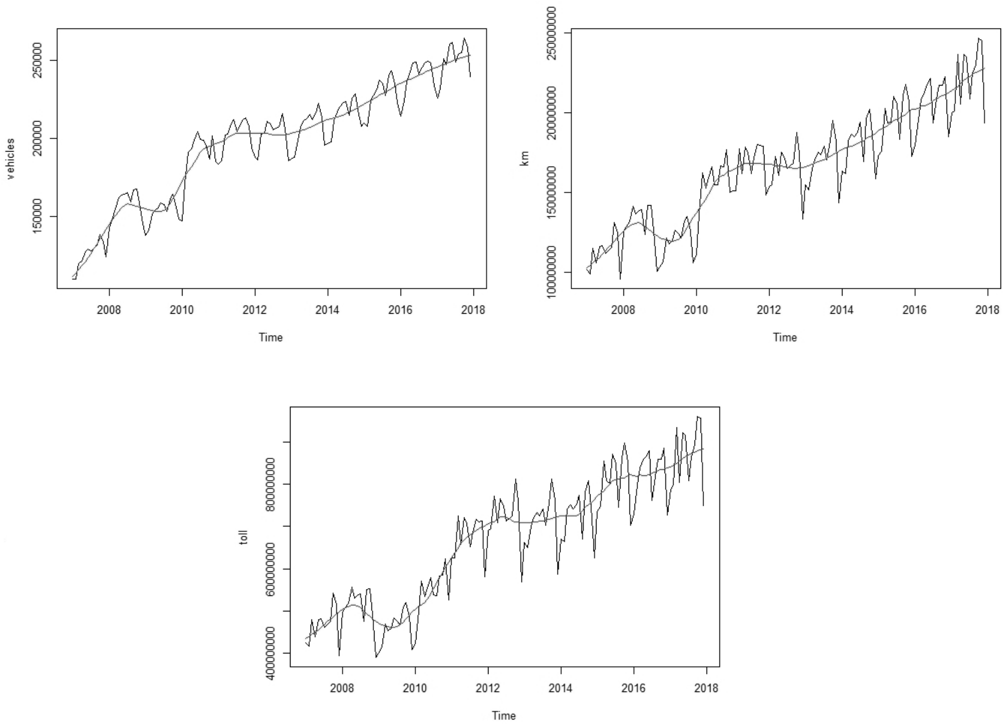

Figure 2.1: Plots of *vehicles*, *km* and *toll*

of exports and imports of goods and services, is used as another dependent variable - *trade*. All these three variables - *rgdp*, *ngdp* and *trade* - are published by Czech Statistical Office (2018) as quarterly time series and they are adjusted from seasonal and calendar effects. As in a case of toll data, the monotonic trends of variables *rgdp*, *ngdp* and *trade* are identified. Cox-Stuart test[3] was used with the alternative hypothesis that upward or downward trend is present. P-values are indeed equal to zero; thus the null hypothesis is rejected, and it is concluded that variables *rgdp*, *ngdp* and *trade* indeed follow the trend.

This thesis follows the study of Döhrn (2013), who used month over month variation of seasonal and working day adjusted industrial production as a dependent variable. Czech Statistical Office (2018) provides basic industrial production index that is adjusted for seasonal and calendar effects. It has a quarterly frequency with the average month of 2015 which equals to 100, and

[3]Results from Cox-Stuart test can be seen in Table D1 in Appendix D.

Variable	Frequency	Unit	Description
rgdp	Quarterly	Million CZK	Real gross domestic product of the Czech Republic, current prices, seasonally adjusted
ngdp	Quarterly	Million CZK	Nominal gross domestic product of the Czech Republic, current prices, seasonally adjusted
trade	Quarterly	Million CZK	Foreign trade of the Czech Republic, external balance of goods and services, seasonally adjusted
product	Quarterly	Index	Basic industrial production index of the Czech Republic, seasonally adjusted, average month of 2015 = 100
vehicles	Monthly	Quantity	Number of all heavy-goods vehicles within the Czech Republic that have travelled more than 0 km in a given month
km	Monthly	km	Total kilometers travelled by all heavy-goods vehicles within the Czech Republic
toll	Monthly	CZK	Total toll collected from all heavy-goods vehicles within the Czech Republic

Table 2.1: Description of variables

it includes total industry as the sum of mining and quarrying, manufacturing, and production and distribution of electricity, gas and air conditioning. Using Cox-Stuart test[4], the null hypothesis cannot be rejected[5] and it is concluded that variable *product* does not follow monotonic trend. Table 2.1 summarizes the description of all variables used in the thesis.

2.2 Seasonal adjustment of toll data

The aim of this section is to remove seasonal and calendar effects, which may in other way disfigure the accurate movement of given variables later used in the thesis. Time series can generally be divided into four components - trend, cycle, seasonal and irregular. The irregular component is the residual after removing trend, cycle and seasonal components, and it shows the random fluctuations of the series (Jiann, 2005).

Seasonality was tested only for toll data, i.e. for variables *vehicles*, *km* and *toll*, because dependent variables were already seasonally and calendar adjusted. Non-parametric Friedman test, which attempts to examine whether months of the year affect the number of vehicles, the total kilometers travelled and total toll collected, respectively, is applied. The test compares the months' variances, which are assumed to be caused by a seasonal factor, with the residual variances that are considered to be derived from the irregular component. It analyses whether the variations within the year are regular and repetitive enough to be dependably classified as the seasonal movements (Eurostat, 2018b).

The 12 samples are derived from seasonal-irregular ratios, i.e., the estimates of detrended series, of size $n_1, n_2, ..., n_{12}$, where each lower index denotes a different level of seasonality. It is assumed that, firstly, each of the 12 samples is derived from random variable X_j which follows the normal distribution with mean equal to m_j and standard deviation σ. Secondly, the seasonality has a effect only on the means of the distribution, not on their

[4]Results from Cox-Stuart test can be seen in Table D1 in Appendix D.

[5]The $\alpha = 0.05$ is used.

variances. Then the following null hypothesis of no seasonality is tested against the alternative:

$$H_0 : m_1 = m_2 = ... = m_{12}$$

$$H_A : m_p \neq m_q, \text{ for one pair of } p, q \text{ at least}$$

In case of all three variables (*vehicles*, *km* and *toll*), the null hypothesis is rejected with p-values indeed equal to 0. It is confirmed that the months affect given variables and the seasonality is present. Therefore there is a need to remove seasonal component which was detected. Let m be a seasonal period, where $m = 4$ for quarterly data and $m = 12$ for monthly data, and values of m are called seasonal indices.

In R software[6], which is used in the thesis research, trend and business-cycle component are taken as one. Therefore, later on in this thesis trend component will denote both trend and cycle components. Let X_t, S_t and I_t denote non-seasonally adjusted time series, trend component, seasonal component and irregular component, respectively, at the period t. It is assumed that the magnitude of seasonal and irregular variations of time series used in the thesis does not depend on the size of the trend component; thus decomposition model is considered to be additive, which can be expressed as:

$$X_t = T_t + S_t + I_t \tag{2.1}$$

This section looks for the estimation of seasonal component, which is then removed, leaving behind trend and irregular components. This thesis follows Hyndman and Athanasopoulos (2014) method of classical seasonal decomposition applying additive model with a seasonal period of $m = 12$. It is assumed that seasonal components are constant from year to year. Firstly, a moving average method is used to estimate the trend. A moving average of order m ($m - MA$) is defined as:

[6] The function *decompose* is used from package *stats*.

$$\hat{T}_t = \frac{1}{m}\sum_{j=-k}^{k} X_{t+j} \tag{2.2}$$

where $m = 2k + 1$ for odd m and $m = 2k$ for even m. Averaging values of time series within k period of time t gives the estimate of the trende component T at the time t. This average removes some of the unpredictability in the data and lets a smooth trend component. On one hand, if seasonal period m is odd number, then $m - MA$ is used to estimate trend, however, on the other hand, when m is even number, then $2 \times m - MA$ is applied, which is just a moving average to a moving average, i.e., a moving average of order m is taken, and then another moving average of order 2 to the results is applied. In case of toll data used in this thesis with a seasonal period of $m = 12$, a moving average of order 2×12 is expressed as:

$$\hat{T}_t = \frac{1}{2}\left[\frac{1}{12}\sum_{j=-6}^{6} X_{t-j}\right] \tag{2.3}$$

Secondly, detrended series is calculated as $X_t - \hat{T}_t$. Averaging the detrended values for each month gives the estimate of the seasonal component for that month. For instance, the average of all detrended values of January in the data provides the seasonal index for January. Seasonal indices for all months are adjusted to secure that their sum is equal to zero and string together to obtain the estimate of seasonal component $\hat{S}_t$. Finally, the remaining (irregular) component is calculated by subtracting the both estimated trend and seasonal components:

$$\hat{I}_t = X_t - \hat{T}_t - \hat{S}_t \tag{2.4}$$

Seasonal adjusted data removes the seasonal component, leaving behind the trend and irregular components $\hat{T}_t + \hat{I}_t$. Figure 2.2 illustrates the seasonally adjusted toll data (dark grey plot) and non-seasonally adjusted data (grey plot).

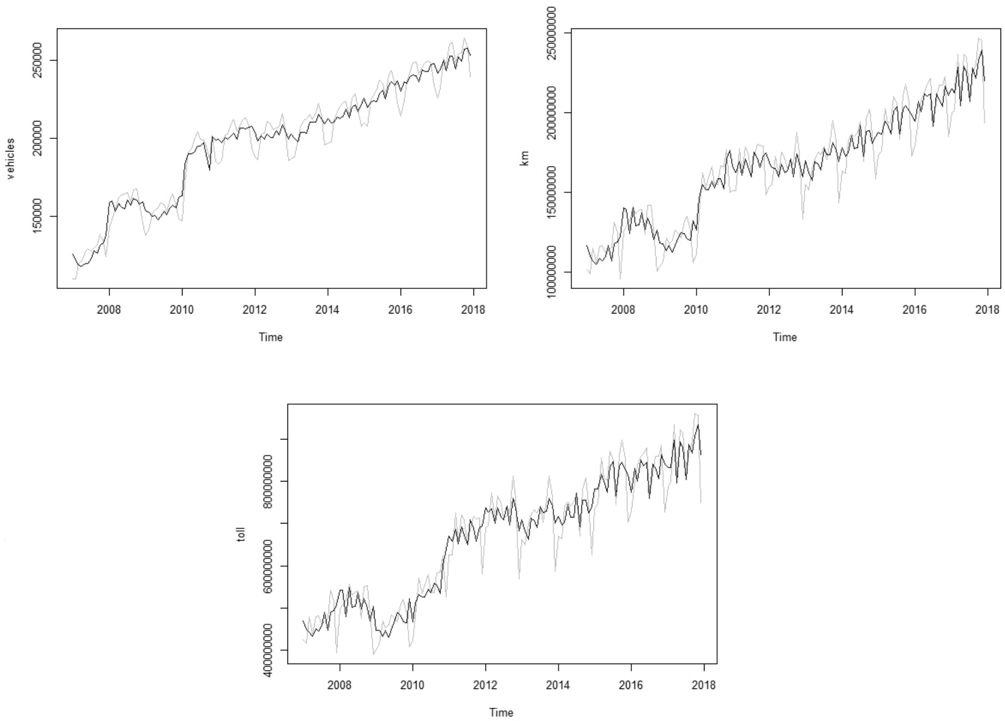

Figure 2.2: Seasonal adjusted *vehicles*, *km* and *toll*

2.3 Stationarity

Stationary time series is defined as one whose distribution, in particular, its variance, mean, or timewise covariance do not change over time (Wooldridge, 2015). The assumption of stationarity is considered to be essential in analyzing time series because non-stationary time series cannot be used in regression analysis since it may generate spurious regression, i.e., a misleading relationship in which variables are not causally linked to each other. However, if two or more non-stationary time series are cointegrated, which means that stationary linear combination of these variables exists, they can be still used in regression modeling (Hyndman and Athanasopoulos, 2014).

Thus it is important to examine whether given time series are stationary or not. The way to do that is to use a unit root test. The one applied in this thesis is Augmented Dickey-Fuller test. Multiple regression is defined as

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_k y_{t-k} \tag{2.5}$$

where k is the number of lags included in the regression. Then the first-order differencing is applied, i.e., $\Delta y_t = y_t - y_{t-1}$:

$$\Delta y_t = \theta y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + ... + \beta_k \Delta y_{t-k} \tag{2.6}$$

where $\theta = \beta_0 - 1$. Dickey-Fuller test is then carried out under the null hypothesis of non-stationarity.

$$H_0 : \theta = 0, \text{ or } \beta_0 = 1, \text{ non-stationary}$$
$$H_A : \theta < 0, \text{ or } \beta_0 < 1, \text{ stationary}$$

Value of the test statistics is computed as

$$DF_T = \frac{\hat{\theta}}{SE(\hat{\theta})} \tag{2.7}$$

and then compared with the relevant critical value[7] for Dickey-Fuller test. If test statistic is lower than the critical value, the null hypothesis is rejected; thus time series is stationary. Table 2.2 provides test statistics and p-values for all seasonally adjusted variables used in the thesis. It can be seen that all variables are non-stationary.

At the end of this section, all these variables are transformed to stationary. To do that, the percentage change between quarters is applied.

Variable	Dickey-Fuller test statistic	P-value	Stationarity
vehicles	-3.1622	0.09728	Non-stationary
km	-2.4404	0.3934	Non-stationary
toll	-2.0459	0.5574	Non-stationary
rgdp	-0.66374	0.9659	Non-stationary
ngdp	-0.64679	0.9674	Non-stationary
trade	-1.5611	0.7475	Non-stationary
product	-2.5829	0.3430	Non-Stationary

Table 2.2: Summary of Dickey-Fuller tests for variables

[7]In the thesis $\alpha = 0.05$ is used.

However, in case of toll data, adjustment have to be done first. Seasonally adjusted toll data - *vehicles*, *km* and *toll* - are converted from monthly data to quarterly. The reason for that is due to the availability of dependent variables as quarterly time series. In case of *km* and *toll*, it is done by summation of given three consecutive months into one quarter. The quarter data are obtained which can be then transformed to stationary by applying the percentage change. However, variable *vehicles* cannot be transformed the same way. This variable contains the information about the number of vehicles which travelled more than zero kilometers in given month. The summation can cause that same vehicle might be included up to three times in given quarter, which might distort results. Thus, it is assumed that percentage change of the number of vehicles is the sum of changes between given three consecutive months. For instance, the percentage change of the number of vehicles in the first quarter is the sum of percentage change between January and February, and between February and March. This might prevent the problem. Nevertheless, this is just assumption; therefore, results should be assessed with caution. Percentage changes of all variables are obtained which make the variables stationary. Thus, they can be used in further analysis.

Chapter 3

Methodology

In this chapter, theoretical methods which are used in the thesis research are provided. The aim of this thesis is to estimate several types of models for each dependent variable that will be later on compared to each other. The most appropriate model for each dependent variable will be then selected using both Akaike and Bayesian information criteria. Firstly, dynamic models with suitable statistical tests are presented and the efficiency of these models to predict dependent variables in sample period is demonstrated. Secondly, theoretical background about ARIMA models is presented. Finally, regressions with ARIMA errors, where appropriate errors structures are selected using Hyndman et al. (2007) algorithm, are provided. Regressions with ARIMA errors are then compared with ARIMA models to identify whether additional information derived from toll data is relevant.

3.1 Dynamic models

In this section, first five models, which will be later on compared, are provided. Dynamic models describes the behaviour of the examined object over time. These models are generally used in cases when objects's behaviour is best presented as a set of the states occurred in specified sequence. In this thesis, distributed-lag model, that is a type of dynamic model in which the effect of regressor on regressand occurs over time preferably than all at the same time, is applied. An equation of distributed-lag model is defined as

$$y_t = \alpha_0 + \sum_{s=0}^{\infty} \beta_s x_{t-s} + \epsilon_t \tag{3.1}$$

where ϵ_t is an error term (Wooldridge, 2015).

Firsts three models are regarded to be a finite distributed-lag models of order 1, which are a special case of equation 3.1. It is implemented for all combinations of dependent and independent variables, i.e. regression of the real GDP on the number of kilometers travelled, regression of real GDP on the number of vehicles, etc. The general model can be therefore specified as

$$\Delta y_t = \alpha + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \epsilon_t \tag{3.2}$$

where y is dependent variable ($rgdp$, $ngdp$, $trade$ and $product$, gradually), x is independent variable ($vehicles$, km and $toll$, gradually) and ϵ_t is an error term.

The set of these OLS regressions is run, and the hypothesis that β_k (for $k = 0, 1, 2, 3, 4$) significantly differs from zero is tested. If it is the case for at least some independent variable, it can indicate a significant effect of independent variables derived from toll data on dependent variables representing indicators of economic activity. According to the work of Döhrn (2013), it is supposed that quarterly differences in kilometers travelled will have a stronger effect on variation in industrial production than quarterly differences in toll collected and a number of vehicles.

Now, all three independent variables are included in regressions, however it need to be check whether multicollinearity is not present. Table 3.1 shows correlation between independent variables. Multicollinearity is

	Δkm	$\Delta vehicles$	$\Delta toll$
Δkm	1.00	0.51	0.64
$\Delta vehicles$	0.51	1.00	0.19
$\Delta toll$	0.64	0.19	1.00

Table 3.1: Correlation matrix of variables derived from toll data

present whenever an independent variable is highly correlated with one or more other independent variables in the equation (Allen, 1997). The highest correlation from given correlation matrix is between the number of kilometers travelled and the amount of toll collected, which is equal to 0.64. This correlation is not too large; therefore it does not look like there is a problem with multicollinearity and all three variables can be included.

Another two models including all three variables are selected, with one lag and two lags, respectively. The reason for choosing these two models is that it can be interesting to identify whether change in independent variables not only quarter ago, but also a half year ago, has significant effect on economic activity. Then two sets of OLS regressions are run.

$$\begin{aligned}\Delta y_t =&\alpha + \beta_0 \Delta km_t + \beta_1 \Delta km_{t-1} + \beta_2 \Delta vehicles_t \\ &+ \beta_3 \Delta vehicles_{t-1} + \beta_4 \Delta toll_t + \beta_5 \Delta toll_{t-1} + \epsilon_t\end{aligned} \tag{3.3}$$

$$\begin{aligned}\Delta y_t =&\alpha + \beta_0 \Delta km_t + \beta_1 \Delta km_{t-1} + \beta_2 \Delta km_{t-2} + \beta_3 \Delta vehicles_t + \beta_4 \Delta vehicles_{t-1} \\ &+ \beta_5 \Delta vehicles_{t-2} + \beta_6 \Delta toll_t + \beta_7 \Delta toll_{t-1} + \beta_8 \Delta toll_{t-2} + \epsilon_t\end{aligned} \tag{3.4}$$

where y is dependent variable (*rgdp*, *ngdp*, *trade* and *product*, gradually) and ϵ_t is an error term.

Equations 3.3 and 3.4 represent finite distributed-lag models of order 1 and 2, respectively, with all three dependent variables. Furthermore, the hypothesis whether β_k (where $k = 0, 1, 2, ..., 5$ for model 3.3 and $k = 0, 1, 2, ..., 8$ for model 3.4) differs from zero is tested. It will show which variable has a significant effect on the prediction of economic activity and which not. .

Three models were obtained. To compare them, both Akaike and Bayesian information criteria are used. Akaike information criterion[1] (AIC) and Bayesian information criterion[2] (BIC) are applied to estimate the relative

[1] Akaike information criterion was defined by Japanese statistician Hirotugu Akaike in 1974.
[2] Bayesian information criterion was introduced by Gideon E. Schwarz in 1978.

quality of given models relatively to each other. Akaike information criterion is based on the in-sample fit of the model to estimate the likelihood to predict future values (Tran and Arabnia, 2015). On the other hand, Bayesian information criterion is based on likelihood function and computes the trade-off between complexity of the model and its fit. The values of AIC and BIC of models with the same dependent variable are then compared. Model with a lower AIC or BIC among all models predicting the same dependent variable is regarded to have a better fit and be preferable than others. Akaike information criterion and Bayesian information criterion, respectively, are estimated by following equations:

$$AIC = 2k - 2ln(\hat{L}) \tag{3.5}$$

$$BIC = ln(n)k - 2ln(\hat{L}) \tag{3.6}$$

where $\hat{L}$ is a maximized value of the likelihood function for the model defined as $L(\theta|x) = P(x|\theta)$, θ is a set of parameters, x is a given observed data, n is the number of observations in x and k is the number of parameters estimated by the model. Due to these criterion, not only it can be seen which regressor is preferable in forecasting economic activity and likely should be included in the regression, but also it can be found out which model predicting given dependent variable is the most appropriate. However, these criteria do not say anything about the absolute quality of models. For instance, if all models are poor, it will only choose the best model from poor ones (Burnham and Anderson, 2003).

3.2 ARIMA models

This thesis follows Döhrn (2013). He used AR(1) (autoregressive model of order 1), and MA(1) (moving-average model of order 1) errors in regressions identifying the effect of transportation data, especially kilometers travelled, on the variation of industrial production. In this section, ARIMA models for each dependent variable are estimated. The main reason is that ARIMA models will be later on compared with regression with ARIMA errors.

Whether the value of both AIC and BIC of regression with ARIMA errors is lower than for ARIMA model, then information about truck transport is relevant and independent variables derived from toll data should be included in regression.

ARIMA (AutoRegressive Integrated Moving Average) models are considered to be a class of time series models which are used to understand properties of time series and to forecast their behaviour in the future. ARIMA model consists of three parts, i.e., autoregressive (AR), interactive (I) and moving-average (MA). Autoregressive model, in other words, means regression of the variable itself. It uses a linear combination of past values of the forecast variable to predict this variable; thus autoregressive model of order p, $AR(p)$, can be expressed by the following equation:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + e_t \tag{3.7}$$

where c is a constant, $\phi_1, ..., \phi_p$ are parameters and e_t is white noise. It can be seen that equation 3.7 is quite similar to the equation 3.2, however, lagged values of y_t are used as independent variables rather than another variable x_t (Hyndman and Athanasopoulos, 2014).

On the other hand, the moving-average model uses past forecast errors in a regression instead of past values of the forecast variable y_t. Moving-average model of order q, $MA(q)$, is regarded to be defined by the following equation:

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} \tag{3.8}$$

where c is a constant, e_t is white noise and $\theta_1, ..., \theta_q$ are parameters. Nevertheless, the values e_t are unobservable; thus equation 3.8 is not a regression in a regular way. Last part of ARIMA model is called integrated part of the model, $I(d)$, and it is present in case of non-stationary data. The differencing is applied, and integrated part provides degree of first differencing involved (d) to eliminate non-stationarity (Hyndman and Athanasopoulos, 2014).

By a combination of differencing with autoregressive and moving-average models, AutoRegressive Integrated Moving Average (ARIMA) model is obtained. Full $ARIMA(p, d, q)$ model can be expressed by the following equation:

$$y'_t = c + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + ... + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} + e_t \quad (3.9)$$

where y'_t is differenced series, c is a constant and e_t is white noise. Right side of equation 3.9 contains both lagged values of y_t - $AR(p)$ - and lagged forecast errors e_t - $MA(q)$ - with degree of first differencing equal to d.

ARIMA model is constructed for all four dependent variables. Hyndman et al. (2007) algorithm is used to fit the best ARIMA model of series. This algorithm combines unit root test (especially Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test) and minimizing maximum likelihood estimation (MLE) and the small-sample-size corrected version of Akaike information criterion (AICc). AICc is corrected AIC for smaller sample size expressed by following formula:

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1} \quad (3.10)$$

where n stands for the size of sample and k denotes the number of parameters.

Hyndman et al. (2007) algorithm follows four steps. Firstly, the degree of first differencing d is detected by repeating KPSS tests. Secondly, time series is differenced by d times and AICc are minimized to obtain values of p and q. Four models - $ARIMA(2, d, 2)$, $ARIMA(0, d, 0)$, $ARIMA(1, d, 0)$ and $ARIMA(0, d, 1)$ - are given and the model with the lowest AICc is selected. Whether d is equal to zero, then constant c is included in the model, however, whether $d \geq 1$ then the constant c is set to equal to zero. A model which is called "current model" is obtained. Thirdly, variations on current model are considered by to ways: (i) p and/or q of the current model is varied by ± 1, (ii) constant c is included or excluded from the current model. The model with the smallest AICc so far, which can be current model or one of its variation, is selected and it turns into the new current model. Finally, the third step is repeated as long as no smaller AICc can be found.

By Hyndman et al. (2007) algorithm, also seasonal ARIMA model can be selected. This model is formed not only as non-seasonal ARIMA model, but it also includes addition seasonal term. Thus it can be written as

$$ARIMA(p, d, q)(P, D, Q)_m$$

where (p, d, q) represents non-seasonal part of the model, (P, D, Q) denotes seasonal part of the model and m equal to the number of periods per season ($m = 12$ per monthly data, $m = 4$ per quarterly data). For instance, $ARIMA(1, 1, 1)(1, 1, 1)_m$ model is defined as:

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)y_t = (1 + \theta_1 B)(1 + \Theta_1 B^4)e_t \quad (3.11)$$

Non-seasonal terms of the model are multiplied by seasonal terms and it gives the equation 3.11, where $(1 - \phi_1 B)$, $(1 - B)$, $(1 + \theta_1 B)$ depict non-seasonal $AR(1)$, non-seasonal difference, non-seasonal $MA(1)$, respectively, and $(1 - \Phi_1 B^4)$, $(1 - B^4)$, $(1 + \Theta_1 B^4)$ represent seasonal $AR(1)$, seasonal difference, seasonal $MA(1)$, respectively (Hyndman and Athanasopoulos, 2014). B is regarded to be the backward shift operator which, in case of this thesis, means:

$$By_t = y_{t-1}$$

$$Be_t = e_{t-1}$$

After application of algorithm by Hyndman et al. (2007), four ARIMA models are chosen for each dependent variable. These ARIMA models are shown in Table 3.2. It can be seen, that only for variable $trade$, non-seasonal ARIMA model is considered, otherwise seasonal ARIMA model is chosen with four periods per the season. Moreover, the hypotheses whether each coefficient differ from zero are tested for all parameters in the equation 3.11.

Variable	Preferable ARIMA model
$\Delta rgdp$	$ARIMA(1,0,0)(1,0,0)_4$ with non-zero mean
$\Delta ngdp$	$ARIMA(0,1,1)(0,0,1)_4$
$\Delta trade$	$ARIMA(0,0,1)$ with non-zero mean
$\Delta product$	$ARIMA(1,0,0)(2,0,0)_4$ with zero mean

Table 3.2: List of peferable ARIMA models using Hyndman et al. (2007) algorithm

3.3 Regression with ARIMA error

On the one hand, multiple regression models gives a lot of relevant information from independent variables, but they do not provide dynamics of the dependent variable as can be done by ARIMA models. On the other hand, ARIMA models produce an amount of information from the past observations of time series, but they do not include other appropriate information from additional independent variables. Thus, these two models are combined to obtain regression with ARIMA errors (Hyndman and Athanasopoulos, 2014). In this section, regression with ARIMA error are estimated and they represent the last models.

Let n_t is error series following ARIMA model. For instance, if error series n_t follows ARIMA(1,1,1), then the regression with ARIMA(1,1,1) error can be expressed as:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_{1,t} + ... + \beta_k x_{k,t} + n_t, \\ (1-\phi_1 B)(1-B)n_t &= (1+\theta_1 B)e_t \end{aligned} \tag{3.12}$$

where e_t is white noise.

Regression with ARIMA errors is run for all four dependent variables. Equation 3.13 represents general regression with ARIMA error for dependent variables used in the thesis, where n_t is ARIMA error and c is a constant.

$$\Delta y_t = \beta_1 \Delta vehicles_t + \beta_2 \Delta km_t + \beta_3 \Delta toll_t + n_t \tag{3.13}$$

To estimate equation 3.13, n_t need to be calculated at first. However, it cannot be done without estimation of coefficients β_1, β_2, β_3 and structure of ARIMA error terms. Thus Hyndman et al. (2007) algorithm is applied to specify ARIMA error terms. Table 3.3 summarized this specification. Then coefficients can be estimated. Furthermore, hypotheses whether each coefficient in equation 3.13, where y_t is given dependent variable and n_t is ARIMA error following the structure provided by Table 3.3, are tested.

Dependent variable	ARIMA error
$\Delta rgdp$	$ARIMA(1,0,0)(1,1,0)_4$
$\Delta ngdp$	$ARIMA(0,1,1)$
$\Delta trade$	$ARIMA(1,0,0)$
$\Delta product$	$ARIMA(1,1,1)(0,0,1)_4$

Table 3.3: List of ARIMA errors structures for regressions with ARIMA errors

Both AIC and BIC are computed for each model. AIC of ARIMA model and regression with ARIMA error for each dependent variable are compared to detect whether the additional information provided by variables derived from toll data are relevant. If value of AIC for regression with ARIMA error is lower than the value of AIC for ARIMA model, variables containing information about truck transport should be included.

Overall, seven models - five dynamic OLS regressions, ARIMA models and regressions with ARIMA error - are obtained. Values of AIC and BIC are computed for each model and models predicting the same dependent variable are compared. The model with the lowest value of both AIC and BIC is supposed to be the most appropriate to predict the given variable representing economic activity.

Chapter 4

Results

This chapter presents the results of the thesis research. Seven models are obtained which are then compared. Firstly, the chapter focuses on a set of finite distributed-lag models of order 1 for all combinations of dependent and independent variables. Secondly, regressions with all three predictors with both 1 and 2 lags are presented. Thirdly, it provides results from ARIMA models and regression with ARIMA errors. These two models are compared to see whether the additional information derived from toll data is relevant in predicting given dependent variable. Finally, a discussion is provided to select the most appropriate model for each dependent variable. This is done by comparing values of both AIC and BIC.

4.1 Finite distributed-lag models of order 1

In this section, first three models are presented. The set of regressions was run, i.e., a finite distributed-lag model of order 1 for all combinations of dependent and independent variables. Presence of heteroskedasticity and autocorrelation was tested for each model using Breusch-Pagan and Breusch-Godfrey tests, respectively. Whether heteroskedasticity or/and autocorrelation was present, then heteroskedasticity and autocorrelation consistent (HAC) errors were estimated. If this correction is not done, standard errors are biased, thus *t statistics* can no longer be used for drawing inferences. In Appendix A, Tables A1 and A2 show regression results for all combination of variables.

It can be seen that variable Δkm is considered to have the best ability to predict economic activity from given independent variables. Table 4.1 provides results for distributed-lag models of order 1 with the independent variable as the percentage change in kilometers travelled for all four dependent variables.

	$\Delta rgdp_t$	$\Delta ngdp_t$	$\Delta trade_t$	$\Delta product_t$
(Intercept)	0.45	0.63^{*}	2.49	-0.39
	(0.26)	(0.28)	(2.57)	(0.67)
Δkm_t	0.17^{***}	0.03	1.30^{*}	0.23^{**}
	(0.03)	(0.04)	(0.61)	(0.07)
Δkm_{t-1}	0.03	0.03	2.99	0.21^{**}
	(0.01)	(0.03)	(1.78)	(0.06)
R^2	0.41	0.04	0.14	0.36
Adj. R^2	0.38	-0.01	0.10	0.32
Num. obs.	42	42	42	42
RMSE	0.89	1.04	37.01	1.96

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

All values in brackets are standard errors.

Table 4.1: Regression results: Distributed-lag models of order 1 with independent variable Δkm

Models with the dependent variable as a percent change in real GDP and percent change in industrial production index are regarded to have the best fit with adjusted R^2 which accounts for 0.38 and 0.32, respectively, and relatively small root-mean-square error[1]. Significant positive effects of percent change in the number of kilometers travelled on the percent change in real GDP, the volume of foreign trade and industrial production index were found. For instance, the 10% increase in the number of kilometers travelled leads to 1.7% increase in real GDP. This effect is significant even at 0.1% significance level. Therefore, toll data can be a useful tool to predict whether there will be an increase or drop in real GDP within given quarter

[1]Ranges of dependent variables are shown in Table C1

before information about GDP will be officially published [2]. However, in case of the model predicting the volume of trade, the 10% increase in the number of kilometers travelled leads to 13% increase in the volume of foreign trade. The effect is significant at 5% significance level and seems to be quite large. Thus there might be some omitted variable bias which might magnify this effect. Furthermore, a positive effect of percent change in the number of kilometers travelled in the previous quarter on the percent change in industrial production index in present quarter was detected. This effect is significant at 1% significance level.

Models with the number of vehicles as an independent variable have generally the worse fit compared to other models with adjusted R^2 no higher than 0.06[3]. This might be caused by assumption that percentage change of the number of vehicles is the sum of changes between given three consecutive months. This way might not be ideal, but it can be challenging to find an adequate way how to deal with this type of problem. Moreover, two significant effects of the number of vehicles were identified. Firstly, the percent change in the number of vehicles affects percent change in real GDP positively. Secondly, the effect of percent change in the number of vehicles in the previous quarter on percent change in industrial production index in the present quarter was found. These effects are significant at 5% significance level.

In case of models with the change in the amount of toll collected as an independent variable, three significant effects of percent change in the amount of toll collected in the previous quarter were detected. These models are shown in Table 4.2. Firstly, the positive effect on the percent change in real GDP was found. This effect is significant at the 5% significance level, and it indicates that 10% increase in the amount of toll collected in given quarter will lead to 0.4% increase in real GDP in the next quarter. Secondly, the positive effect on the percent change in the volume of foreign trade was

[2]Precise data about GDP are lagged by 9 weeks in general (Eurostat, 2018a), while toll data are available within a few weeks after the end of given month (VÝROČENKY.cz, 2018).

[3]The results of these models can be seen in Tables A1 and A2 in Appendix A.

	$\Delta rgdp_t$	$\Delta trade_t$	$\Delta product_t$
(Intercept)	0.59	2.56	−0.43
	(0.31)	(2.41)	(0.66)
$\Delta toll_t$	0.07	1.28	0.21
	(0.07)	(0.90)	(0.10)
$\Delta toll_{t-1}$	0.04*	3.24*	0.27**
	(0.02)	(1.38)	(0.09)
R^2	0.12	0.17	0.45
Adj. R^2	0.07	0.12	0.42
Num. obs.	42	42	42
RMSE	1.09	36.47	1.81

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

All values in brackets are standard errors.

Table 4.2: Regression results: Distributed-lag models of order 1 with independent variable $\Delta toll$

found. It shows that 10% increase in the amount of toll collected in given quarter will lead to 32.4% increase in the volume of foreign trade in the next quarter. Compared to the real GDP, this effect is considered to be relatively large. Thus there might be some omitted variable. This can cause bias which might distort results and make this effect unreliably large. Finally, the positive effect on the percent change in industrial production index was found. This effect is even significant at 1% significance level. The 10% increase in the amount of toll collected in given quarter will lead to the 2.7% increase in industrial production index in the next quarter. It can be seen that the amount of toll collected might be more useful than the number of kilometers travelled because the amount can predict economic activity even earlier.

Nevertheless, none significant effect was found in models predicting nominal GDP. All these models have negative adjusted R^2 with relatively high root-mean-square error compared to the range of the dependent variable. It might be said that variables derived from toll data are not a good predictor of nominal GDP.

4.2 Dynamic models with three independent variables

In this section, another two models for each dependent variable, which will be later on compared with others, are presented. Now, all three independent variables are included in regressions. Table 4.3 shows dynamic models with one lag which includes all three independent variables.

	$\Delta rgdp_t$	$\Delta ngdp_t$	$\Delta product_t$	$\Delta trade_t$
(Intercept)	0.48	0.66^{*}	-0.56	3.28
	(0.26)	(0.28)	(0.56)	(3.59)
Δkm_t	0.22^{***}	0.05	0.23^{***}	-0.14
	(0.04)	(0.03)	(0.05)	(1.16)
Δkm_{t-1}	0.01	0.03	-0.03	3.41
	(0.02)	(0.03)	(0.10)	(1.85)
$\Delta vehicles_t$	-0.12	-0.16	-0.53^{*}	-0.29
	(0.10)	(0.11)	(0.21)	(2.81)
$\Delta vehicles_{t-1}$	0.06	-0.01	0.43	-8.48^{*}
	(0.08)	(0.12)	(0.27)	(3.37)
$\Delta toll_t$	-0.05	0.01	0.09	1.35
	(0.04)	(0.02)	(0.06)	(1.24)
$\Delta toll_{t-1}$	0.02	-0.00	0.29^{***}	1.77
	(0.03)	(0.04)	(0.08)	(1.31)
R^2	0.44	0.07	0.55	0.23
Adj. R^2	0.34	-0.10	0.48	0.10
Num. obs.	42	42	42	42
RMSE	0.92	1.08	1.72	36.91

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

All values in brackets are standard errors.

Table 4.3: Regression results: Dynamic models with one lag

Compared to models with only one independent variable, the fit of the model predicting industrial production index with all three independent variables increased considerably. For this model, adjusted R^2 equals to 0.48. Three significant effects were found. Firstly, a positive effect of the percent change of kilometers travelled was identified, which is significant at the 0.1%

significance level. Secondly, a positive effect of percent change in toll collected in the previous quarter was detected. The 10% increase in the amount of toll collected in given quarter will lead to 2.9% increase in industrial production index. Moreover, this effect is significant at the 0.1% significance level. Finally, the negative effect of percent change in the number of vehicles was found which seems surprising. However, that might occur due to the limitation of data containing information about vehicles. Data contains the number of ALL heavy-goods vehicles which travelled within the territory of Czech Republic regardless of what kind of transport is involved. It consists of transport which imports to and exports from the Czech Republic, on the one hand, and domestic transport and transit, on the other. Thus, there might be an increase in vehicles which do not influence industrial production index. Furthermore, this negative effect is significant at 5% significance level. Thus, using 5% significance level might result in significant type I error.

Similar negative effect of percent change in the number of vehicles was found in the model predicting the volume of foreign trade. Moreover, this effect is relatively large. This can be caused by including too many variables which might lead to overspecification bias. The fit of this model is comparatively same as in case of models with only one independent variable with adjusted R^2 equals to 0.10 and root-mean-square error which equals to 36.91[4].

The positive effect of percentage change in the number of kilometers travelled on percent change in real GDP was identified. This effect is significant at the 0.1% significance level. It indicates that the 10% increase in the number of kilometers travelled leads to 2.2% increase in real GDP. This effect is slightly higher than in case of the model with percent change in the number of kilometers as the only independent variable. The fit of this model slightly decreased with adjusted R^2 of 0.34 and higher root-mean-square error of 0.92.

Even in case of the model with all three independent variables, none

[4]Compared to the range of dependent variable in Table C1.

significant effect on nominal GDP was detected. Adjusted R^2 remained negative with a relatively high root-mean-square error.

Now, another lag is added to the regression with all three independent variables. Table 4.4 shows the results. It can be seen that fit of models worsened compared to dynamic models with one lag. A few significant effects of second lag were detected. However, some other significant effects disappeared. The bad fit, relatively high root-mean-square error and weak significance might be caused by too many variables included in regressions which indicates overspecification bias. This type of models is rather a practical example than an accurate model for predicting economic activity.

4.3 ARIMA models and models with ARIMA error

In this section, last two types of models - ARIMA models and regressions with ARIMA error - are presented. ARIMA models from Table 3.2 were constructed, and their results are shown in Table B1[5]. These ARIMA models are specially presented to be compared with regressions which include ARIMA errors. Whether the value of AIC of regression with ARIMA error is lower than for ARIMA model, then information derived from toll data is relevant and should be included in the regression.

Regressions with ARIMA errors which are summarized in Table 3.3 were run. Results of these regressions are shown in Table 4.5. It can be seen that except model predicting real GDP the value of both AIC and BIC is lower for ARIMA model than for regression with ARIMA error. This means that ARIMA model is preferable and variables derived should not be included in regression containing ARIMA terms. In case of regression predicting real GDP, the significant positive effect of percent change in the number of kilometers travelled was identified. It is significant at 0.1% significance level, and it implied that the 10% increase in the number of kilometers travelled leads to the 1.6% increase in the real GDP.

[5]See Appendix B.

	$\Delta rgdp_t$	$\Delta ngdp_t$	$\Delta product_t$	$\Delta trade_t$
(Intercept)	0.50^{*}	0.67^{*}	-0.57	6.57
	(0.24)	(0.29)	(0.57)	(3.38)
Δkm_t	0.24^{***}	0.08	0.23^{*}	-0.71
	(0.04)	(0.04)	(0.10)	(1.87)
Δkm_{t-1}	-0.03	-0.01	-0.02	4.62
	(0.03)	(0.04)	(0.11)	(3.20)
Δkm_{t-2}	0.07^{*}	0.08	0.01	-2.13
	(0.03)	(0.05)	(0.08)	(2.95)
$\Delta vehicles_t$	-0.19	-0.24	-0.52	1.45
	(0.11)	(0.13)	(0.26)	(2.16)
$\Delta vehicles_{t-1}$	0.15	0.10	0.40	-7.07
	(0.09)	(0.14)	(0.26)	(6.65)
$\Delta vehicles_{t-2}$	-0.22^{*}	-0.26	0.04	5.56
	(0.11)	(0.18)	(0.31)	(8.89)
$\Delta toll_t$	-0.04	0.02	0.09	1.06
	(0.05)	(0.03)	(0.08)	(1.26)
$\Delta toll_{t-1}$	0.03	0.02	0.29^{**}	1.50
	(0.02)	(0.04)	(0.10)	(1.34)
$\Delta toll_{t-2}$	-0.04	-0.04	0.00	-1.84
	(0.02)	(0.03)	(0.05)	(1.41)
R^2	0.48	0.12	0.56	0.35
Adj. R^2	0.32	-0.13	0.43	0.17
Num. obs.	41	41	41	41
RMSE	0.94	1.10	1.82	35.91

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

All values in brackets are standard errors.

Table 4.4: Regression results: Dynamic models with two lags

	$\Delta rgdp_t$	$\Delta ngdp_t$	$\Delta trade_t$	$\Delta product_t$
ar1	0.55^{***}		-0.34^{*}	0.61^{**}
	(0.15)		(0.15)	(0.19)
sar1	-0.59^{***}			
	(0.14)			
$vehicles_t$	0.01	-0.09	1.07	-0.35
	(0.09)	(0.13)	(4.73)	(0.26)
km_t	0.16^{***}	0.03	1.81	0.16
	(0.04)	(0.05)	(1.75)	(0.09)
$toll_t$	-0.04	0.02	1.22	-0.04
	(0.03)	(0.04)	(1.51)	(0.08)
ma1		-0.65^{***}		-0.95^{***}
		(0.17)		(0.12)
sma1				-0.40^{*}
				(0.17)
AIC	106.69	123.77	438.80	183.47
AICc	109.32	125.43	440.42	186.76
BIC	116.67	132.46	447.60	195.63
Log Likelihood	-47.35	-56.88	-214.40	-84.73

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

All values in brackets are standard errors.

Table 4.5: Results: Regression with ARIMA error

4.4 Preferable models

In the last section of this chapter, the values of AIC of models predicting the same dependent variable representing economic activity are compared. The lower value of AIC indicates better fit and the model with the lowest value of AIC is considered to be preferable.

	rgdp	*ngdp*	*trade*	*product*
FDL model with km	114.43	127.45	427.42	180.51
FDL model with vehicles	133.28	128.76	432.72	194.37
FDL model with toll	131.57	127.99	426.18	174.07
Dynamic model with one lag	120.48	134.13	430.65	173.31
Dynamic model with two lags	122.15	134.83	420.53	176.08
ARIMA model	119.82	118.55	435.94	180.42
Regression with ARIMA error	106.69	123.77	438.80	183.47

Table 4.6: Values of AIC for all models

Table 4.6 shows the values of AIC for each model[6]. Firstly, the preferable model for real GDP seems to be a regression with $ARIMA(1,0,0)(1,1,0)_4$ error with the value of AIC which equals to 106.69. The positive effect of percent change in the number of kilometers travelled on the percent change of real GDP was identified. This effect is significant at 0.1% significance level. It shows that 10% increase in the number of kilometers travelled leads to 1.6% increase in real GDP. Secondly, $ARIMA(0,1,1)(0,0,1)_4$ model is selected as the most appropriate model predicting nominal GDP. This indicates that none of three variables containing information derived from toll data is an appropriate predictor of nominal GDP due to additional information from toll data is not relevant. The value of AIC of this model accounts for 118.55, and it can be concluded that nominal GDP cannot be predicted by toll data. The main reason for that might be that toll data do not contain the effect of price level changes. Thirdly, the preferable model predicting the volume of foreign trade is a dynamic model with two lags containing all three variables. In this model, none significant effect

[6] Values of BIC are shown in Table E1 in Appendix E.

was found. That might be another indications that there was some bias in previous models where incredibly large coefficients appeared. Finally, the preferable model for predicting industrial production index seems to be a dynamic model with one lag containing all three variables. The value of AIC accounts for 173.31. Three significant effects were detected. However, the negative effect of percent change in the numbers of vehicles on percent change in industrial production index needs to be taken with caution due to the limitation of data. This effect is significant at 5% significance level that is why it can suffer from significant type I error. Preferable models for each dependent variable are summarized in Table 4.7.

	$\Delta rgdp_t$	$\Delta ngdp_t$	$\Delta trade_t$	$\Delta product_t$
(Intercept)			6.57	-0.56
			(3.38)	(0.56)
ar1	0.55^{***}			
	(0.15)			
sar1	-0.59^{***}			
	(0.14)			
$\Delta vehicles_t$	0.01		1.45	-0.53^{*}
	(0.09)		(2.16)	(0.21)
Δkm_t	0.16^{***}		-0.71	0.23^{***}
	(0.04)		(1.87)	(0.05)
$\Delta toll_t$	-0.04		1.06	0.09
	(0.03)		(1.26)	(0.06)
ma1		-0.59^{***}		
		(0.14)		
sma1		-0.29		
		(0.17)		
Δkm_{t-1}			4.62	-0.03
			(3.20)	(0.10)
Δkm_{t-2}			-2.13	
			(2.95)	
$\Delta vehicles_{t-1}$			-7.07	0.43
			(6.65)	(0.27)
$\Delta vehicles_{t-2}$			5.56	
			(8.89)	
$\Delta toll_{t-1}$			1.50	0.29^{***}
			(1.34)	(0.08)
$\Delta toll_{t-2}$			-1.84	
			(1.41)	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

All values in brackets are standard errors.

Table 4.7: Results: Preferable model for each dependent variable

Conclusion

The thesis assessed the ability of seasonally adjusted toll data to predict Czech economic activity which is represented by four variables - real GDP, nominal GDP, industrial production index and volume of foreign trade. Toll data included the number of kilometers travelled, the number of vehicles travelled within the territory of the Czech Republic and the amount of toll collected, all by heavy-goods vehicles within the territory of the Czech Republic. The aim of this thesis was to construct different types of models for each dependent variable which were then compared to each other using both Akaike and Bayesian information criterion. The model with the lowest value of AIC and BIC was considered to be preferable.

For each dependent variable, a set of seven models was run. Firstly, finite distributed models with one lag, which contain only one independent variable, were constructed. Secondly, dynamic models with one and two lags, which include all three independent variables, were composed. Finally, ARIMA models and regressions with ARIMA errors were constructed and compared. It was observed that only for the model predicting real GDP, additional information derived from toll data is relevant and should be included in the model which contains ARIMA terms.

In sample used, the number of kilometers travelled, and the amount of toll collected are regarded to be good predictors of economic activity. On the other hand, in models containing all three independent variables, the effect of the number of vehicles on economic activity was detected to be significant and opposite than it was expected. It showed that increase in the number of vehicles travelled within the territory of Czech Republic leads to a decrease in real GDP, industrial production index and volume of foreign trade, ceteris paribus. This finding might be caused by the limitation of data. For instance, toll data consists of the number of all vehicles travelled within the territory of the Czech Republic independently on kind of transport included. Thus the negative effect of the number of vehicles might be caused by an increase in transit and domestic transport which do not influence the volume of foreign trade of the Czech Republic. Furthermore, these negative effects are significant at 5% significance level; thus it might result in type I error.

Toll data can predict both industrial production index and real GDP. As the most appropriate model predicting real GDP, regression with ARIMA error was selected. The significant positive effect of percent change on percent change in real GDP was identified. On the other hand, dynamic model with one lag containing all three variables appeared to be preferable for predicting industrial production index. The positive effect of the number of kilometers travelled on industrial production index was found. This finding is consistent with Askitas and Zimmermann (2013), and Döhrn (2013). Furthermore, a significant positive effect of the amount of toll collected in the previous quarter on industrial production index in present quarter was detected. Therefore, it can be said that the amount of toll collected might be more useful than the number of kilometers travelled. Döhrn (2013) mentioned that the advantage of the model predicting industrial index using toll data is not too big because the difference between the publication date of toll data and industrial production index is quite small. Nevertheless, the ability of the amount of toll collected to predict industrial production index one quarter ahead might result in removing issue with the small publication difference.

According to thesis research, it was found out that toll data might not be a good predictor of both the volume of foreign trade and nominal GDP. ARIMA model seems to be the most appropriate model for predicting nominal GDP. It can be concluded that information derived from toll data is not relevant for predicting nominal GDP. On the other hand, dynamic model with two lags containing all three variables was detected as preferable model for predicting the volume of foreign trade. However, none significant effect was found. That might indicate that there was a bias in other models with the volume of foreign trade as a dependent variable where incredibly large coefficients appeared.

To sum it up, firstly, the hypothesis that variables containing information about truck transportation can predict economic activity was approved for real GDP and industrial production index. In case of nominal GDP, this hypothesis was rejected. Secondly, there was found a significant negative effect of the number of vehicles travelled on industrial production index. Thus, the hypothesis that all effects of variables derived from toll data are positive was rejected. Thirdly, at least one significant effect was found in the preferable model predicting real GDP and industrial production index. Thus, this hypothesis was approved for this two dependent variables. In case of nominal GDP and the volume of foreign trade, this hypothesis was rejected. Fourth, the hypothesis that dynamic models are more accurate than ARIMA models was rejected for nominal GDP. In case of the volume of foreign trade and industrial production index, this hypothesis was approved. Finally, the hypothesis that information about truck transportation are relevant and should be included in regression with ARIMA error is approved only for real GDP. For remaining dependent variables, this hypothesis is rejected.

Nevertheless, in few years, time series produced by electronic toll collection system will have longer time spam which can provide more enhanced analysis. One can also deal with monthly data instead of quarterly. Moreover, focusing on the limitation of data described above and its removing seems to be essential for further analysis. For instance, separating transit and domestic transport from the other and finding a new way to convert monthly variable containing information about the number of vehicles to quarterly might lead to more appropriate results.

Bibliography

Allen, M. P. (1997). The problem of multicollinearity. *Understanding regression analysis*.

Aprigliano, V., G. Ardizzi, and L. Monteforte (2017). Using the payment system data to forecast the Italian GDP.

Askitas, N. and K. Zimmermann (2011). The Toll Index: Innovation-based Economic Telemetry.

Askitas, N. and K. Zimmermann (2013). Nowcasting business cycles using toll data. *Journal of Forecasting*.

Aylward, A. and J. Glen (2000). Some international evidence on stock prices as leading indicators of economic activity. *Applied Financial Economics*.

Bodo, G., A. Cividini, and L. F. Signorini (1991). Forecasting the Italian industrial production index in real time. *Journal of Forecasting*.

Broaddus, A. and C. Gertz (2008). Tolling heavy goods vehicles: Ioverview of European practice and lessons from German experience. *Transportation Research Record: Journal of the Transportation Research Board*.

Burnham, K. P. and D. R. Anderson (2003). *Model selection and multimodel inference: A practical information-theoretic approach.* Springer Science & Business Media.

Camba-Mendez, G., G. Kapetanios, R. J. Smith, and M. R. Weale (2001). An automatic leading indicator of economic activity: Forecasting GDP growth for European countries. *The Econometrics Journal*.

Carlsen, M. and P. E. Storgaard (2010). Dankort payments as a timely indicator of retail sales in Denmark. Technical report, Danmarks Nationalbank Working Papers.

Černý, K. (2008). Electronic toll collection in the Czech Republic. In *International Conference, Sofia (Bulgaria)*.

Czech Statistical Office (2018). [Online]. Accessed 2018-04-20. Available on https://www.czso.cz/.

Döhrn, R. (2013). Transportation data as a tool for nowcasting economic activity - The German road pricing system as an example.

Edey, M. L., J. Pleban, et al. (1991). *Indicators of Economic Activity: A Review*. Reserve Bank of Australia.

Estrella, A. and G. A. Hardouvelis (1991). The term structure as a predictor of real economic activity. *The journal of Finance*.

European Comission: MEMO/10/489 (2010). Review of the Directive on charging Heavy Goods Vehicles - Eurovignette Directive- Questions and Answers. [Online]. Accessed 2018-04-06. Available on http://europa.eu/rapid/press-release_MEMO-10-489_en.htm.

European Comission: Mobility and Transport (2018). [Online]. Accessed 2018-04-06. https://ec.europa.eu/transport.

Eurostat (2018a). [Online]. Accessed 2018-04-25. Available on https://ec.europa.eu/eurostat.

Eurostat (2018b). Seasonal adjustment e-learning courses. [Online]. Accessed 2018-04-25. Available on https://ec.europa.eu/eurostat/sa-elearning/take-elearning-course-seasonal-adjustment.

Gill, T., D. Perera, D. Sunner, et al. (2012). Electronic indicators of economic activity. *Reserve Bank of Australia Bulletin*.

Gordon, R. J. (1992). Productivity in the transportation sector. In *Output measurement in the service sectors*. University of Chicago Press.

Guzavicius, A., V. Barkauskas, and V. Tamulis (2013). Nowcasting business cycles using transportation index. In *Poceedings of the 10th International Scientific Conference*.

Han, X. and B. Fang (2000). Four measures of transportation economic importance. *Journal of Transportation and Statistics*.

Hyndman, R. J. and G. Athanasopoulos (2014). *Forecasting: Principles and practice*. OTexts.

Hyndman, R. J., Y. Khandakar, et al. (2007). *Automatic time series for forecasting: The forecast package for R*. Monash University, Department of Econometrics and Business Statistics.

Jiann, H. C. (2005). Seasonal Adjustment of Time Series. *Economic Accounts Division, Singapore Department of Statistics, Statistics Singapore Newsletter*.

Kenny, S. (2017). The role of road charging in improving transport: How road charging supports efforts to decarbonise the sector. *Transport Environment*.

Kokkinen, A. and H. Wouters (2018). Euro area and European Union GDP flash estimates at 30 days.

Lahiri, K., W. Yao, and P. Young (2004). Transportation and the economy: Linkages at business-cycle frequencies. *Transportation Research Record: Journal of the Transportation Research Board*.

McKinnon, A. C. (2006). A review of European truck tolling schemes and assessment of their possible impact on logistics systems. *International Journal of Logistics*.

Ministry of Transport (2018). [Online]. Accessed 2018-02-19. Available on `https://www.mdcr.cz`.

MYTO.cz (2018). [Online]. Accessed 2018-03-14. Available on `http://mytocz.eu/`.

Rosenberg, J. V. and S. Maurer (2008). Signal or noise? Implications of the term premium for recession forecasting.

Rutkowska, A. (2015). Properties of the Cox-Stuart test for trend in application to hydrological series: The simulation study. *Communications in Statistics-Simulation and Computation*.

Sůra, J. (2018). Státní navigace. Propojí data od řidičů s údaji o nehodách či z mýtných bran. [Online]. Accessed 2018-04-13. Available on https://www.denik.cz/ekonomika/statni-navigace-propoji-data-od-ridicu-s-udaji-o-nehodach-ci-z-mytnych-bran-20180404.html.

The National Bureau of Economic Research (2018). [Online]. Accessed 2018-04-25. Available on http://www.nber.org/.

Tran, Q. N. and H. Arabnia (2015). Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology.

Šitner, R. (2017). Mýtný systém vytváří spoustu užitečných dat, stačí je sebrat. [Online]. Accessed 2018-04-06. Available on https://archiv.ihned.cz/c1-65965220-mytny-system-vytvari-spoustu-uzitecnych-dat-staci-je-sebrat.

VÝROČENKY.cz (2018). [Online]. Accessed 2018-02-02. Available on `http://www.vyrocenky.cz/`.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Nelson Education.

Yao, V. W. (2005). The causal linkages between freight and economic fluctuations. *International Journal of Transport Economics/Rivista internazionale di economia dei trasporti*.

List of Tables

List of Figures

Appendix A

	$\Delta rgdp_t$	$\Delta rgdp_t$	$\Delta rgdp_t$	$\Delta ngdp_t$	$\Delta ngdp_t$	$\Delta ngdp_t$
(Intercept)	0.45	0.64	0.59	0.63^{*}	0.74^{*}	0.67^{*}
	(0.26)	(0.42)	(0.31)	(0.28)	(0.29)	(0.28)
Δkm_t	0.17^{***}			0.03		
	(0.03)			(0.04)		
Δkm_{t-1}	0.03			0.03		
	(0.01)			(0.03)		
$\Delta vehicles_t$		0.26^{*}			-0.05	
		(0.12)			(0.13)	
$\Delta vehicles_{t-1}$		0.01			0.04	
		(0.07)			(0.08)	
$\Delta toll_t$			0.07			0.03
			(0.07)			(0.03)
$\Delta toll_{t-1}$			0.04^{*}			0.01
			(0.02)			(0.04)
R^2	0.41	0.08	0.12	0.04	0.00	0.02
Adj. R^2	0.38	0.03	0.07	-0.01	-0.05	-0.03
Num. obs.	42	42	42	42	42	42
RMSE	0.89	1.12	1.09	1.04	1.06	1.05

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

All values in brackets are standard errors.

Table A1: Regression results: Distributed-lag models of order 1 (part 1)

	$\Delta trade_t$	$\Delta trade_t$	$\Delta trade_t$	$\Delta product_t$	$\Delta product_t$	$\Delta product_t$
(Intercept)	2.49	8.48	2.56	-0.39	-0.04	-0.43
	(2.57)	(4.77)	(2.41)	(0.67)	(0.72)	(0.66)
Δkm_t	1.30^{*}			0.23^{**}		
	(0.61)			(0.07)		
Δkm_{t-1}	2.99			0.21^{**}		
	(1.78)			(0.06)		
$\Delta vehicles_t$		4.84			0.34	
		(3.47)			(0.28)	
$\Delta vehicles_{t-1}$		-1.22			0.49^{*}	
		(1.76)			(0.19)	
$\Delta toll_t$			1.28			0.21
			(0.90)			(0.10)
$\Delta toll_{t-1}$			3.24^{*}			0.27^{**}
			(1.38)			(0.09)
R^2	0.14	0.02	0.17	0.36	0.10	0.45
Adj. R^2	0.10	-0.03	0.12	0.32	0.06	0.42
Num. obs.	42	42	42	42	42	42
RMSE	37.01	39.42	36.47	1.96	2.31	1.81

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

All values in brackets are standard errors.

Table A2: Regression results: Distributed-lag models of order 1 (part 2)

Appendix B

	$\Delta rgdp_t$	$\Delta ngdp_t$	$\Delta trade_t$	$\Delta product_t$
ar1	0.57^{***}			0.62^{***}
	(0.12)			(0.12)
sar1	-0.32^{*}			-0.33^{*}
	(0.15)			(0.15)
intercept	0.76		10.26	
	(0.23)		(3.54)	
ma1		-0.59^{***}	-0.36^{*}	
		(0.14)	(0.14)	
sma1		-0.29		
		(0.17)		
sar2				-0.31
				(0.18)
AIC	119.82	118.55	435.94	180.42
AICc	120.88	119.18	436.55	181.47
BIC	126.87	123.76	441.22	187.46
Log Likelihood	-55.91	-56.27	-214.97	-86.21

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

All values in brackets are standard errors.

Table B1: Results: ARIMA models

Appendix C

Variable	Minimum	Maximum
$\Delta rgdp$	-2.337102	3.358488
$\Delta ngdp$	-2.101726	2.962682
$\Delta trade$	-69.04978	172.97892
$\Delta product$	-8.131353	4.786734

Table C1: Ranges of dependent variables

	km	toll	vehicles
Min.:	95,857,517	389,284,675	110,218
1st Qu.:	136,367,983	532,497,801	164,375
Median:	168,805,278	702,360,621	205,821
Mean:	166,073,642	666,033,022	197,678
3rd Qu.:	193,403,363	773,615,183	224,366
Max.:	246,520,999	961,631,079	264,353

Table C2: Summary statistics for toll data

Appendix D

Variable	Num. obs.	Statistics	p-value
km	66	66	$p < 0.001$
vehicles	66	66	$p < 0.001$
toll	66	65	$p < 0.001$
rgdp	22	22	$p < 0.001$
ngdp	22	22	$p < 0.001$
trade	22	22	$p < 0.001$
product	22	16	0.052

Statistics provides the number of pairs with a signal "+"

Table D1: Results: Cox-Stuart test

Appendix E

	rgdp	*ngdp*	*trade*	*product*
FDL model with km	121.38	134.40	434.37	187.46
FDL model with vehicles	140.23	135.72	439.67	201.32
FDL model with toll	138.52	134.94	433.13	181.02
Dynamic model with one lag	134.38	148.03	444.55	187.21
Dynamic model with two lags	141.00	153.68	439.38	194.93
ARIMA model	126.87	123.76	441.22	187.46
Regression with ARIMA error	116.67	132.46	447.60	195.63

Table E1: Values of BIC for all models