CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

High Frequency Price Index of Construction Materials: Nowcasting of Producer Prices

Master's thesis

Author: Bc. Josef Štefl
Study program: Economics and Finance
Supervisor: Mgr. Petr Polák, M.Sc., Ph.D.
Year of defense: 2024

Josef Štefl
71688324@fsv.cuni.cz

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, July 28, 2024

....

Josef Štefl

Acknowledgments

The author would like to thank the thesis supervisor Mgr. Petr Polák, M.Sc., Ph.D., M.Sc. for his time, supportive guidance, and helpful feedbacks. The author would also like to express his heartfelt gratitude to Tereza for her unwavering willingness to discuss the issues that arose while working on this thesis and for her ability to find solutions. Finally, the author extends his gratitude to the employees not only of Czech power plants, whose cooperation was indispensable to this research, and to the Royal Society for the Prevention of Birds.

Bibliographic record

Štefl, Josef. *High Frequency Price Index of Construction Materials: Nowcasting of Producer Prices. Master's thesis*. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2024, Advisor: Mgr. Petr Polák, M.Sc., Ph.D., M.Sc.

Abstract

The thesis explores the potential of leveraging *big data* in economics through the authors' own weekly *Price Index of Construction Materials* (CMPI) derived from web-scraped data. This area of research was selected due to the relatively scant literature on nowcasting producer prices, coupled with the ready availability of data concerning building materials from online eshops. The research, conducted between October 2021 and June 2024, focuses on three primary objectives: enhancing the accuracy of market development descriptions through high-frequency data, exploring the relationship between inflation in construction materials and commodity prices, and nowcasting the official construction material index using the web-scraped CMPI. The findings confirm the feasibility of measuring building materials inflation over the long term using web-scraped data. It turns out that in a normal period free from the impacts of pandemics or war conflicts, commodity prices do not significantly contribute to estimating the CMPI. Conversely, the nowcasting of the official index one month ahead can be substantially enhanced by incorporating web-scraped data, demonstrating its practical utility in economic forecasting.

JEL classification C43, C55, C80, E31, E37

Keywords inflation, high frequency price index, big data, web scraping, producer prices, nowcasting

Title High Frequency Price Index of Construction Materials: Nowcasting of Producer Prices

Abstrakt

Práce zkoumá potenciál využití *velkých dat* v ekonomii prostřednictvím vlastního týdenního *indexu cen stavebních materiálů* (CMPI) odvozeného z dat získaných metodou *web scraping*. Tato oblast byla vybrána kvůli relativní stručnosti literatury o prognózování cen výrobců, kdy jsou ale data týkající se stavebních materiálů snadno dostupná z eshopů. Výzkum prováděný v období od října 2021 do června 2024 se zaměřuje na tři hlavní cíle: zvýšení přesnosti popisu vývoje trhu prostřednictvím vysokofrekvenčních dat, zkoumání vztahu mezi inflací u stavebních materiálů a cenami komodit a *nowcasting* oficiálního indexu stavebních materiálů s využitím dat CMPI získaných z internetu. Zjištění potvrzují proveditelnost měření inflace stavebních materiálů v dlouhodobém horizontu pomocí dat z webu. Ukazuje se, že v normálním období bez dopadů pandemie nebo válečného konfliktu ceny komodit k odhadu indexu CMPI významně nepřispívají. Naopak nowcasting oficiálního cenového indexu na jeden měsíc dopředu může být podstatně vylepšen začleněním online dat, což dokazuje jejich praktickou využitelnost v ekonomické prognóze.

Klíčová slova inflace, vysokofrekvenční cenový index, velká data, web scraping, ceny výrobců, prognózování

Název práce Vysokofrekvenční cenový index stavebních materiálů: Prognózování cen výrobců

Contents

List of Tables

List of Figures

Acronyms

ADF	Augmented Dickey-Fuller (test)
ACF	AutoCorrelation Function
AIC	Akaike Information Criterion
AICc	Corrected Akaike Information Criterion
ARIMA	Autoregressive Integrated Moving Average (model)
ARIMAX	Autoregressive Integrated Moving Average (model) with Exogenous Inputs
BIC	Bayesian Information Criterion
BLS	Bureau of Labor Statistics
CPI	Consumer Price Index
CZK	Czech Crown
CZSO	Czech Statistical Office
CNB	Czech National Bank
CMPI	Price Index of Construction Materials / Construction Materials Price Index
CWPI	Construction Work Price Index
DM	Diebold-Mariano (test)
DFM	Dynamic Factor Model
DOM	Document Object Model
DPI	Digital Price Index
eCPI	Consumer Price Index from online data
eCPIDB	eCPI Database
GAS	Generalized Autoregressive Score (model)
GARCH	Generalized AutoRegressive Conditional Heteroskedasticity (model)

GDFM	Generalized Dynamic Factor Model
GDP	Gross Domestic Product
HICP	Harmonised Index of Consumer Prices
HTML	HyperText Markup Language
Lasso	Least Absolute Shrinkage and Selection Operator
MA	Moving Average
MIDAS	Mixed Data Sampling (model)
MIPI	Price Index of Material Inputs for Construction Works
OLS	Ordinary Least Squares
ONS	Office for National Statistics
PACF	Partial AutoCorrelation Function
PCE	Personal Consumption Expenditures
pcs	pieces
p.p.	percentage point
PPI	Producer Price Index
RMSE	Root Mean Square Error
SARMA	Seasonal Autoregressive Integrated Moving Average (model)
UK	United Kingdom
URS	Institute of Rationalization of Construction Work
VAR	Vector Autoregression Model
VECM	Vector Error Correction Model
WTI	West Texas Intermediate (crude oil)

Chapter 1

Introduction

Changes in the price level are captured by movements in inflation. This central measure of macroeconomics is usually expressed using two complementary dimensions that cover both the demand and supply sides of the market – consumer prices and producer prices. For its correct quantitative description, the national statistical offices publish the respective consumer price indices (CPIs) and producer price indices (PPIs). The former shows changes in the prices of goods and services purchased by households and the latter measures changes in domestic production input prices. (Wei, 2018)

The Czech Statistical Office (CZSO) divides producer prices into four different categories and constructs the corresponding indices: Industrial Producer Price Index, *Construction Work Price Index* (CWPI), Market Services Price Index, and Agricultural Producer Price Index. Although the methodology of these statistical measurements differs, at least one thing remains common. All of them are, to a greater or lesser extent, based on statistical questionnaires sent out to selected manufacturers and service providers with a frequency not higher than monthly. (CZSO, 2023)

In 2021, we commenced the research focused on producer prices data collection in the construction work market segment (Štefl, 2022). To this end, we created a *Construction Materials Price Index* (CMPI) that captures weekly movements of price level of building material inputs – a crucial determinant of final construction prices.

The underlying data was gathered using a method called *web scraping*, which is essentially the automated extraction of information from public websites. For three quarters, until the middle of 2022, our web-scraping robots were browsing the full offers of the three major Czech eshops with building materials every week. The data collected in this way allowed us to conclude that a reliable price index could be constructed using this method.

After the completion of this first stage of the research described in the aforementioned bachelor's thesis (Štefl, 2022), the data collection continued until 30 June 2024. The resulting dataset of 144 weekly records covers the 11 quarters, i.e., almost 3 years of the volatile inflationary environment in the Czech republic capturing the covid-19 recovery period and the impacts of the Russian invasion of Ukraine in 2022, including the subsequent energy crisis.

The first but rather ancillary objective of this thesis is to validate the results of previous research, i.e., to assess the relevance of CMPI in the long term. For this purpose, we compare the index with other relevant indices published by the CZSO (chiefly with CWPI and the complementary MIPI – *Price Index of Material Inputs for Construction Work*) and look in more detail at possible differences and background influences. In this section, among other things, we describe the evolution of producers' inflation in this Central European market with weekly granularity for the first time, but also study the impact of the change in MIPI weighting system from the onset of 2023. This assessment becomes particularly interesting given that the MIPI, in addition to being part of producer prices, also influences the general CPI through its integration with imputed rents. (CZSO, 2024)

Assuming that the official average annual inflation of construction materials measured by the MIPI was double digit for the consecutive two years (10.9% in 2021, 20.7% in 2022) and then it started approaching the Czech National Bank (CNB) 2% target during 2023 (3.8%), this part alone (see Section 4.1) provides an impressive insight into the behaviour of a developed open market economy in times of distress (CZSO, 2024). Moreover, it provides a practical description of the advantages and disadvantages of an alternative method of measuring inflation that potentially refines and accelerates both economic research and economic policy implementation.

The second objective of this scientific endeavour is to develop a reliable and coherent time-series based econometric model to predict the CMPI and hence estimate future changes in the price level of construction materials. Along with the development of the model, an explanation of the methods used and their suitability, this part of the work thoroughly examines the unique dataset in order to evaluate the appropriateness of the econometric approach chosen.

Given the overall volatility and implied non-stationarity of the data over the period under review, real-time data appear to be crucial for the correct determination of the valid econometric model (Groen, 2013). Hence, additional scientific research on the use of high-frequency web-scraped data contributes to further exploring the possible improvements in macroeconomic modelling. In Section 4.2, we propose the models predominantly based on autoregressive moving average (ARIMA) approach.

Similar to inflation, the CNB's key interest rate was also moving dynamically in 2021-2024 period. From a level of 1.5% at the beginning of data collection in the 3rd quarter of 2021, it peaked at 7% in the middle of 2023 and declined to 4.75% by the end of this research in 1st quarter of 2024 (CNB, 2024). The weekly frequency of our dataset enables us to evaluate not only the impact of the interest rate response but extend the ARIMA model of other exogenous variables. In addition to interest rates, the thesis examines the impact of commodity prices, such as steel, copper, crude oil or electricity, and explores the possibilities of using ARIMAX model for predicting CMPI.

Finally, the CMPI itself is used as an external regressor in the ARIMAX model to nowcast the official MIPI. This represents the key objective of our study. Since the CMPI data are available approximately one month before the MIPI, this allows us to nowcast official building materials inflation with considerable accuracy (see Section 4.3).

Many of the methods and analyses presented here would have been impossible to perform before the enormous advances in computing technology in recent years, as a lot of details remained hidden for decades due to the data collection techniques of the time. However, the use of *big data* in economics offers the opportunity to move from the study of statistical samples to entire underlying populations.

This thesis comprehensively summarises three years of practical research and suggests a possible direction for future macroeconomic data collection. In particular, it describes an innovative yet reliable way of nowcasting producer prices.

Chapter 2

Literature review

Notwithstanding the relatively recent emergence of big data applications, specifically web scraping, in economic discourse, scientific efforts in this area have developed at a modest but decent pace in recent years. Conversely, the corpus of the scientific literature that deals in detail with inflation forecasting is already substantial for some time. This section guides the reader through the relevant scholarly texts, starting with the existing economic experience with web scraping methods and ending with econometric methodologies for estimation of inflation that have already been used and are relevant when using high-frequency web-scraped data.

2.1 Web-scraped data in economics

In existing economic research, the applications of online or web-scraped data typically encompass three primary areas: the measurement of inflation, the forecasting (including nowcasting) of inflation, and the exploration of micro price-setting mechanisms. Our work is focusing on the first two mentioned fields.

One of the pioneering studies in this domain was conducted by Lunnemann & Wintr (2006), focusing on price stickiness. They gathered price quotes from the internet, aiming to evaluate the disparities in price rigidity between traditional *brick-and-mortar* shops and online stores across Europe and the USA. The data collection, which spanned one year commencing in December 2004 and involved downloading daily price quotes from online price comparison websites in five different countries, resulted in a core dataset of more than 5 million records. Although the article elucidated pricing behaviour on the internet and its comparison with traditional shops, it did not specifically focus on inflation.

Further progress in this area was achieved through the *Billion Prices Project*, which was launched at MIT in 2008. The initiative of Cavallo & Rigobon (2016) represents a major contribution in terms of the volume of data collected, which is still unprecedented even today. The paper particularly showcases the use of online prices to construct daily price indices across multiple countries and how this approach can help circumvent measurement biases typically associated with traditional methods. The main thrust was on the use of online prices to enhance the accuracy of inflation measurement, which is a critical component in understanding and potentially forecasting inflation, but the paper's specified goal was more aligned with measurement rather than direct forecasting.

Another steps of research on web-scraped data in economics were conducted at the UK Office for National Statistics (ONS) since 2014. Breton et al. (2015) selected three major supermarkets present in both offline and online domain, which is in line with Cavallo & Rigobon (2016), from whose websites they started to download data on a daily basis. The ONS carried out this project as part of a wider initiative called the *Big Data Project*, which aims to innovate and exploit new data sources for statistical purposes. The paper from 2015 details the use of web scraping techniques to collect price data from internet retail sources, a practice that is becoming increasingly viable as internet retailing develops.

In a follow-up update, Breton et al. (2016) streamline the data processing and discuss the progress made in using machine learning techniques for classification and cleaning, an area for development identified in previous papers. Specifically, it explores the application of supervised and unsupervised machine learning techniques to data extracted from web pages and the use of imputation as a means to reduce the negative impact of missing prices in the data.

The ONS *Big Data Project* produced several other working papers focused on web-scraped price indices in which the authors study e.g. clustering of large datasets into price indices (Metcalfe et al., 2016; Payne, 2017; Hayes, 2017), imputing of web-scraped prices (Mayhew, 2016) or the occurance of extreme price changes (Beeson, 2015). Most of these works are based on the dataset collected in the original research by Breton et al. (2015) and focus on either food, clothing or other items commonly included in the CPI.

Powell et al. (2017), in their 14-month research involving the offerings of three large UK supermarkets, discuss the use of online price data to predict the CPI and related statistics more frequently than the previously common approach. Highlighting the disorderly nature of the web-scraped data and the difficulty of reconciling them with standard CPI forecasting methods, the study develops a new model that describes the joint time evolution of latent daily log-inflation measures from internet prices and prices from official surveys and adjusts for different product categories. The authors uncover different dynamic behaviors and levels of predictability across product categories, with a particular focus on aggregation, and demonstrate the ability to accurately forecast category-specific CPI immediately prior to their official launch.

Online indices compiled by *PriceStats*, a spin-off of the *Billion Prices Project,* were used by Bertolotto & Aparicio (2020) in their comparative research on forecasting inflation using online prices in ten developed economies. Although the authors did not collect the data themselves, they describe the data collection methodology extensively and overall demonstrate the robust predictive power of online price data in forecasting inflation, suggesting that it is a valuable tool for economic forecasting beyond traditional methods.

Goolsbee & Klenow (2018) investigate the impact of e-commerce on inflation rates using data from *Adobe Analytics* covering 2014-2017. This platform is another attractive and very worthy alternative to direct web scraping, covering billions of online transactions across hundreds of online retailers and products, similar to the *Billion Prices Project* database. This dataset is particularly rich and provides detailed information on prices, categories and quantities. This latter detail is not usually available directly from websites, which almost never include information on overall attractiveness or number of purchases made, making this dataset more akin to *scanner data*[*)] than web-scraped data.

The additional sales volume information allowed Goolsbee & Klenow (2018) to analyze inflation trends in the e-commerce sector with new insights, so they could focus with their *Digital Price Index* (DPI) on, among other things, the impact of new goods on inflation (i.e., new items in the CPI consumer basket). The so-called *new goods bias* refers to the tendency of traditional inflation measures to be overstated when they fail to fully account for the entry of new products into the market and their impact on consumer choice and spending patterns (Feenstra, 1994). Goolsbee & Klenow (2018) consider this as one possible explanation for their finding that DPI inflation is significantly lower than the traditional consumer price index – 1.3 p.p. per year lower for the categories being compared.

Only a year after the launch of the *Billion Prices Project*, the National Bank of Poland (NBP) launched in 2009 its own research on the consumer price index obtained by web scraping, which it refers to as the *eCPI* (Macias & Stelmasiak, 2019). The study conducts a long-term pseudo-real-time nowcasting experiment using online food and soft drink prices collected from major online retailers in Poland on a highly disaggregated framework, based on which the authors evaluate the effectiveness of web-scraped data in improving nowcasting of Polish food inflation.

*) *Scanner data* is generated by instantly recording the details of items directly at the point of sale and needs to be coordinated with the retailer (Štefl, 2022).

In their research paper, Macias & Stelmasiak (2019) emphasize the importance of proper selection and classification of scraped products to ensure that the data is usable and accurate for forecasting and other applications. The project, originally focused on groceries, expanded in 2017 to include other categories like apparel, footwear, home-improvement equipment, drugs, or even airplane tickets.

With work summarized by Macias et al. (2022) showing that online price estimates effectively predict food inflation and that incorporating these prices into a simple, optimized model further improves the accuracy of the prediction. The approach presented in the paper outperforms various other methods, including judgmental forecasts and traditional benchmarks, especially after the covid pandemic.

In addition to using web-scraped data as their empirical base, the research papers summarized above have one more thing in common: they focus on consumer prices. Regardless of whether they examine the prices of food, clothing, electronics, or even airline tickets, as far as we are aware, no one has yet explored the use of web-scraped data to construct producer price indices.

2.1.1 Data collection

The *Billion Prices* project discusses the methodology of collecting online prices using web scraping technologies in detail and among other facts it highlights the selection of retailers and products to ensure representativeness. The authors stick to direct extraction from the retailers website, so that they do not rely on third parties such as price comparators as it was in the case of Lunnemann & Wintr (2006) study.

On the other hand, for data collection purposes, they rely on third-party web-scraping robot providers *that are able to extract information from any other webpage with a similar structure, storing the information in a database*, so they do not process the code themselves. Consequently, they identify the risk of reliable performance of web-scraping robots in the long run. (Cavallo & Rigobon, 2016)

As opposed to the previous research, Breton et al. (2015) developed their own web scrapers, programmed in Python using the *Scrapy* module, which were set to automatically collect prices for 35 items from the Consumer Prices Index (CPI) basket every day. The Python environment, however, according to Breton et al. (2015), was not prepared for the increasing use of JavaScript, and therefore their web scrapers late in the research only downloaded a limited amount of data from web pages under review.

In the early stages of our research, we ran into a similar problem with JavaScript as Breton (2015), but found that with enough computing power, it is now possible to scrape websites that use it. However, the websites we examined use standard HTML. (Štefl, 2022)

Powell et al. (2017) gathered daily prices for 33 product categories from the websites of three large supermarkets operating in the UK over approximately a 14-month period. On average, they found that a particular item was available for 146 days, studying mostly groceries. One particular problem they faced when collecting data was the inability to reliably distinguish between missing items, as it was not always clear whether the items were missing due to out-of-stocks or a failure of the web scraping robots. The authors also investigated whether the eshops in focus charged the same prices regardless of geographic location, as they claimed, which appeared to be true.

Research by Bertolotto & Aparicio (2020) and Goolsbee & Klenow (2018) both rely on third-party data providers. Although Bertolotto & Aparicio (2020) include a detailed description of the data collection process, a few of their claims turned out to be missleading for a practical research, such as that *the retailer assigns a unique identifier (ID) to each good, which is constant over time*. Our experience shows that the eshop's internal item ID can hardly serve as a unique and consistent identifier (Štefl, 2022). On the other hand, this does not diminish the value of the findings they made, as these problems were simply outsourced, i.e., resolved by another party.

Originally, Macias & Stelmasiak (2019) retrieved data weekly since 2009, but expanded the frequency to daily in 2017. They emphasize computing efficiency in the data retrieval process, so they combine both a *live browsing* approach and *web page parsing*. The first technique renders an action similar to a real-world user – a web-scraping robot directly browses a web page and interacts with its document object model (DOM). This method can also be used to scrape JavaScript web pages, which was unsuccessfully tackled by Breton (2015). In contrast, the web page parsing interacts with the domain only once – when a particular website is loaded, it downloads the entire HTML code and then processes it offline. However, it requires having all the necessary information already recorded in HTML, which is usually not the case with JavaScript. (Zhao, 2017)

In our research endeavor, we address the issues of data collection by developing and deploying custom web scraping robots (see Section 3.1.1). This *in-house* solution based on the HTML parsing enables a complex understanding of the potential limitations and challenges associated with the data collection methodologies we use.

2.1.2 Data processing

The *Billion Prices Project* further defines an efficient way of high-frequency data processing in three general steps, including a brief description of their technological aspects: cleaning, standardization and classification of individual records. Assuming its importance in this field, it is unsurprising that subsequent endeavors have adopted this approach.

Breton et al. (2015) described in detail the scope and scale of data collection, mentioning the wider range of types and locations of prices surveyed compared to traditional CPI collection methods. For example, over 6,000 whisky price quotes per month were collected from websites, reflecting a variety of types such as single malt whiskies or blends. In contrast, for products such as bananas, only seven prices per day were collected from each supermarket. Web scrapers collected three key variables: product name, price, and information about a discount, although the effect of discounts was not the focus of this paper.

Similar details were also focused on by Cavallo & Rigobon (2016), whose robots scraped the product identification number, name, short description, brand, package size, category information, price, and selling price when available, which are then used to correctly classify the data.

Macias & Stelmasiak (2019) emphasize the importance of data quality, as it is not difficult to achieve quantity through web scraping – in their research they obtained over 159 million quotes for 650,000 products. They fruther identify two main discrepancies that may arise between the online price index and official CPI at the micro price level: product classification error and product selection bias.

The first occurs when a product is incorrectly assigned to the wrong group (in the case of the CMPI, this would be the assignment of, for example, *nails* to the *bricks* product group), leading to a mismatch of prices and an incorrect assignment of expenditure weights in the index. (Macias et al., 2022)

The product selection bias tends to be more subtle and results from a mismatch of product types within the correct product group. Thus, its importance in the case of Macias et al. (2022) only becomes apparent when trying to compare the traditional CPI with the eCPI. The bias arises when the variety selected for the online dataset differs from the variety tracked by the statistical office, even if it is correctly classified.

On the practical side of data processing, by far the most detailed treatment of the topic is given in Macias & Stelmasiak (2019). They use a semi-distributed database structure in two forms – *Data Lake* and *Time-Frequency Reduced Relational Database* (eCPIDB). Raw web-scraped data is stored in a data lake (database with no SQL language) on daily basis. The data are then processed into monthly average prices and pulled into the eCPIDB, which is designed for more convenient macroeconomic analysis. The eCPIDB is well structured and allows easy and fast access, selection and classification of data. Its organised and uniform data structure facilitates forecasting exercises and other macroeconomic applications which are then performed by Macias et al. (2022). The structure of their eCPI system reflects the balance between accommodating large volumes of continuously incoming records while maintaining the highest possible data quality.

2.1.3 Advantages of web-scraped data

Additionally, Cavallo & Rigobon (2016) provide an overview of the benefits and drawbacks of web-scraped data. Among other advantages, the authors mention the low cost and possible high frequency of the near real-time data collection process, which we mention first because we consider this attribute to be key.

Cavallo & Rigobon (2016) also reported that the daily frequency made it easier for them to spot errors in the data. Our research confirms that detecting the presence of errors is easy and potentially less impactful, as there are always enough substitute records that can be used to impute missing prices where appropriate (see Section 3.1.2). On the other hand, given a large amount of data, the workload for error handling becomes challenging. Also, the high frequency – even weekly as in our case – puts a lot of demands on the speed of repair since if an error occurs on the side of the web scraping robots and the repair does not take place within a week, one loses two records in a row.

Another important positive feature associated with web-scraped data is that prices are recorded from the moment a product is first offered to consumers until the end of its availability in the store, not from the moment an item in the basket disappears from stores as in traditional data collection. Nevertheless, according to Powell et al. (2017), this is only a theoretical advantage, because in practice it is not easy to distinguish between a real product disappearance and a technical error of a robot performing web scraping. (Cavallo & Rigobon, 2016)

The positive factor that Cavallo & Rigobon (2016) identify in the third place is related to downloading – i.e., tracking – the full offerings of selected retailers. They point out that this results in a much higher level of data cross-sectionality, so this enhanced granularity helps to simplify any quality adjustments. This point is closely related to the possibility of expanding the number of items in the consumption basket compared to the traditional approach, and thus shifting from examining samples to underlying populations (Breton et al. 2016).

One advantage of using web scraping to collect economic data that is not useful for our work, but is worth noting for the comprehensiveness of this thesis, is the ability to execute it remotely (Cavallo & Rigobon, 2016). To illustrate, one can analyse the data for e.g. Israel and Taiwan from Prague in this way and be sure that the same methodology is used in both cases.

The main disadvantage of data collected in this way, which Cavallo & Rigobon (2016) cite as the second, is the lack of information on quantity sold, which in turn is the biggest advantage of scanner data. In this case, all products are treated equally regardless of the quantities sold, unlike traditional methods where price collectors select representative items (Breton et al., 2016). We are reluctant to consider this feature of web-scraped data as a disadvantage in general, as this problem can always be overcome by simply selecting the same items as in the official consumption basket. In this case, the only difference in our view is the origin of the individual quotations. In the case of the CMPI, we have used both approaches and for the traditional consumption basket we have selected materials relevant for the construction of an average family house, thus ensuring that our findings are truly representative (see Section 3.1.2).

An undeniable, but not unsolvable problem remains a characteristic that is common to all big data, namely the high complexity of cleaning and sorting such data. The volume of records extracted in a raw manner usually makes it impossible to use standard cleaning techniques at all. To illustrate, it is undoubtedly easier to properly clean and rank 500 items in a traditional consumer basket than to do the same for all 50,000 items offered by a given eshop. (Breton et al., 2016; Macias & Stelmasiak, 2017)

2.1.4 Imputation, unit price and chained daily indices

One of the quite often mentioned characteristics of high-frequency data obtained by web scraping is that items suddenly appear and disappear in the eshop offer (Cavallo & Rigobon, 2016; Breton, 2015; Cavallo, 2015; Powell, 2017). This problem, if not properly addressed, causes potential inconsistencies in the data and affects the reliability of the resulting price indices (Mayhew, 2016).

An option of preventing this problem often used in the standard CPI construction is called *cell-relative imputation*. This procedure assumes that the price change between the original last missing record for a given item is the same as the average price change of all similar items over the same period. (BLS, 2023; CZSO, 2023)

On the other hand, a simulation study by Mayhew (2016) shows that even simply rolling over the last known price level for a given item minimizes potential imputation bias. This finding is justified in the same study by the fact that the price of an item changes on average every 120 days.

Breton (2015) finds two other methods of constructing price indices that do not require imputation and can be used with a large amount of underlying data, but at the expense of being able to effectively compare the resulting indices with the traditional CPI based on a representative consumption basket.

The *unit price index* is a fixed base index that calculates the average price over a certain period (weekly or monthly). Its consumer basket includes only products with a unit price available in all periods, i.e., items that are absent at least once are completely omitted from the underlying data, which significantly limits the number of prices that can be used to calculate the index. Consequently, this model means that the index may be subject to future revisions as more data become available and products are included and excluded from the sample. In the initial phase of our research, which covered the turbulent period after the Russian invasion of Ukraine, we found this method to be impractical for long-term research when the economy is going through a period of distress, as more than 6% of the offered items we monitored disappeared from week to week. (Breton, 2015; Štefl, 2022)

In contrast, the *chained daily index*[*)] is constructed from bilateral daily indices that are chained into a continuous index. It may include a high number of prices because products need to be sampled only for each pair of consecutive days. This method allows for more extensive use of the data but typically involves more variability because it continuously incorporates new daily price information. Although this type of index may cover the entire offer of the eshop, theoretically even the entire market, its explanatory value for practice, similarly to the unit price index, may not be as high, because at the core both of these types completely lack information on actual expenditure. (Breton, 2015)

Our approach to these challenges is that we compile several indices based on slightly different methodology inspired by both the classical method based on fixed consumer basket and the aforementioned alternatives. We also propose a new procedure that uses adjusted consumer basket for the chained weekly index (see Section 3.1.2).

2.2 Inflation forecasting

In 2023 alone, approximately 18,000 academic articles were published on inflation forecasting, representing an impressive rate of approximately two articles per hour. It would be impossible to attempt to encompass the entire scientific literature in this area in the following pages. Our objective is therefore more modest, namely to explore approaches to inflation forecasting that use high-frequency data and relate them to the specifics of our dataset.

Forecasting inflation is not straightforward in principle. Canova (2007) highlights that forecasting inflation a year ahead proves challenging as most models struggle to outperform a random walk forecast, and some models even perform worse. By and large, his work provides a comprehensive guide to macroeconomic modelling in general and describes a wide range of appropriate models including their properties.

*) From now on, we will refer to this term as *chained weekly* because we work with a weekly frequency in our research.

The aforementioned difficulty of inflation forecasts is further supported by the empirical work of Atkeson & Ohanian (2001), who also find that the naive forecast – assuming that inflation next year will be the same as last year – is remarkably difficult to surpass. Despite using advanced econometric models based on the Phillips curve, their results show that these models do not outperform the simpler naive forecast.

Subsequent analyses reveal that since the mid-1980s, the accuracy of forecasts using traditional multivariate models has not improved significantly over univariate time series models (Fisher et al., 2002). This can be attributed to changes in the behaviour of inflation, which has become less sensitive to factors traditionally used in forecasting models such as Phillips curve (Stock & Watson, 2006).

Faust & Wright (2013) further underscore the debates about the utility of these traditional economic relationships in forecasting. In their study, in addition to comprehensively examining various forecasting models and methodologies, they report the important finding that subjective forecasts often outperform model-based forecasts, especially along certain dimensions. By this they mean forecasts made by experts *who have access to econometric models but add expert judgment to them*.

Nevertheless, it is not only the Phillips curve approach that can be used to forecast inflation. A study by Boughton and Branson (1990) explores the predictive power of commodity prices for forecasting inflation. They find that, historically, commodity prices can often be considered as predictors of changes in inflationary pressures due to their immediate response to supply and demand shocks. However, they note that the link between these prices and actual inflation can be complex and influenced by various global economic conditions. Their conclusions suggest commodity prices can serve as useful indicators and have some predictive power, but do not consistently outperform other forecasting models.

A substantial portion of the research papers focuses on forecasting inflation across short- and medium-term horizons. Given the high frequency of our data, we can narrow the scope of this literature review directly to inflation nowcasting, where relatively fewer academic texts are available. And as Faust & Wright (2013) suggest, accurate nowcasts are crucial for effective forecasting, because they establish the initial conditions from which all future forecasts are projected. In other words, if the starting point (i.e., the nowcast) is flawed, it can lead to compounding errors in short and medium term.

The concept of inflation nowcasting is relatively newer compared to national accounts nowcasting. Angelini et al. (2008) focus on nowcasting and forecasting GDP and its components using a dynamic factor model for real-time applications. This model integrates both high-frequency (monthly) and low-frequency (quarterly) data to provide timely estimates of GDP.

Pseudo real-time forecasting exercise for GDP nowcasting demonstrate that the models based on high-frequency data can outperform several benchmarks, including naive forecasts and bridge equations. Although the *high frequency* in the pioneering works usually means monthly data, it has been able to increase the accuracy and timeliness of current GDP estimates. (Giannone et al., 2007; Angelini et al., 2008)

Monteforte & Moretti (2009) nowcast inflation using a large dataset including multiple variables on daily frequency. As explanatory variables, they used the monthly core inflation index and daily prices of commodities and financial assets, including financial derivatives (e.g. HICP futures contracts). They point out that the inclusion of daily financial data significantly improves the accuracy of inflation forecasting compared to models that use only monthly data.

The use of daily data in inflation estimation can improve not only the performance of the model at the current horizon (nowcasting), but also increase its forecasting accuracy up to one year ahead. The most notable improvements were in the forecasts of the most volatile variables, especially the energy component of inflation, which is often affected by high-frequency price changes. (Modugno, 2011)

As shown by Knotek & Zaman (2015), it is not necessary to use an exhaustive number of variables to increase model accuracy. To inform their nowcasting model, they use a relatively small set of data series. Such a resulting model frequently outperforms the estimates of professional forecasters and consensus surveys, especially when more data is available over the course of a month.

In this respect, the use of web-scraped data is essential, as it is available almost immediately after it is collected and thus provides an instant view of the market. In their comprehensive study, Bertolotto & Aparicio (2020) show that forecasts using online price indices consistently outperform forecasts based on traditional methods, including professional surveys such as the Bloomberg survey of forecasters. They conclude that online indices often predict changes in the consumer price index and provide an early signal of changes in inflation trends, which could revolutionize traditional approaches by incorporating high-frequency, real-time data into economic models.

Macias et al. (2022) also report positive findings on the use of web-scraped data for nowcasting the official CPI. Their framework outperforms various other approaches, including traditional and judgmental methods, and they find that the accuracy of nowcasts increases with the volume of online data. In addition, the prediction quality of their model improved substantially compared to other methods during the covid-19 pandemic, demonstrating the robustness and adaptability of their approach at a time when traditional data collection methods can be disrupted.

2.2.1 CMPI estimation

In our research, we conduct two different forecasting excercies. Before estimating the official index using the CMPI (see Section 4.3), we first estimate the CMPI itself. To this end, we use three different forecasting approaches: autoregression (i.e., using itself), through exogenous variables (i.e., commodity prices), and also by synthesizing both ways (see Section 4.2).

Univariate forecasting

The use of the models based on the autoregression is a cornerstone of statistical time series analysis and forecasting (Canova, 2007). We build our autoregressive models on the work of Meyler et al. (1998), which examines in detail the use of autoregressive integrated moving average (ARIMA) models to forecast inflation in Ireland. The authors concluded that while ARIMA models are useful for short-term inflation forecasting and generally perform well relative to more complex models, they are inherently backward-looking and may not effectively predict inflation turning points or changes in inflation. One should add that Meyler et al. (1998) do not focus on producer prices but use consumer prices captured in the HICP.

In contrast, Tang et al. (2019) primarily focus on estimating producer prices, but their research uses the ARIMA model only as a complement. The authors formulates a sophisticated hybrid forecasting model integrating fuzzy information granulation with genetic algorithm-support vector regression and ARIMA. They consider this method to be particularly useful for handling non-linear relationships in data, while ARIMA, in their view, effectively handles linear aspects and also provides a comprehensive forecasting tool.

In our research, in addition to the ARIMA model, we develop competing multivariate ARIMAX model (an ARIMA model with exogenous variables) incorporating commodity prices, and then compare these approaches. Similarly, we use ARIMAX in the subsequent nowcasting of the official MIPI with CMPI as the external regressor.

Multivariate forecasting

The link between commodity prices and inflation has been confirmed to a greater or lesser extent in many academic papers since the 1980s (Bosworth & Lawrence, 1982; Beckerman & Jenkinson, 1986; Durand & Blöndal, 1988). Although these papers examine data with monthly frequency and focus primarily on consumer prices, many of the methods used and findings are relevant to our research.

Part of our methodological approach (e.g., assigning weights to explanatory variables) is similar to the framework put forward by Boughton & Brenson (1990), who examine the potential of commodity prices to predict inflation in G7 countries. Compared to the authors, however, we contribute by working with a significantly higher frequency of data, and we also focus on estimating producer prices, where the link to commodity prices is more straightforward than for consumer prices.

The work of Boughton and Brenson (1990), which uses a dynamic model in which commodities are treated as either final goods or inputs, confirms the qualitative relationship between commodity prices and consumer price inflation, with commodity price cycles often preceding consumer price cycles. However, the authors generally find the quantitative links between commodity prices and inflation weak and volatile, which reduces the reliability of commodity prices as accurate predictors of inflation.

A more recent study by Hobijn (2008) examines the impact of commodity price fluctuations on US inflation, focusing on the personal consumption expenditure (PCE) index. This analysis covers the period from June 2006 to June 2008, which is shorter than our research time frame. Interestingly, this period was characterized by significant increases in commodity prices, especially grain and oil, which is also true for the time period we analyze (i.e., October 2021 to June 2024). The paper does not deal specifically with inflation forecasting, but concludes that while commodity prices do affect some sectors of the economy, their impact on broader measures of inflation, such as the core PCE index, is limited. Nevertheless, the methodology of quantifying the impact of individual commodities on price level changes to assess the importance of commodity prices in relation to producer prices is useful for our paper.

Ajmera et al. (2012) extend previous research by Hobijn (2008) to a wider range of commodities and a longer time frame. Again, their work does not focus on inflation estimation, but confirms earlier findings, namely that while commodity prices significantly affect some sectors of the CPI, their impact on the overall inflation rate is limited. These papers further motivate our investigation of the impact of commodity prices on the price level in the producer market, which can be seen as a certain segment of the overall market.

To the best of our knowledge, the use of commodity prices as explanatory variables for inflation with higher than monthly frequency has so far been examined only by Modugno (2011). In the area of econometric modeling, our methodology for estimating the CMPI via commodity prices is based on the nowcasting and forecasting techniques defined in this paper for inflation analysis.

Modugno (2011) includes weekly and daily data such as weekly Eurozone oil price statistics and weekly US retail gasoline and diesel prices, along with daily commodity prices on world markets. In this paper, a dynamic factor model is proposed to handle large amount of exogenous variables. This model captures latent factors from observable data that represent underlying economic trends and cycles affecting inflation. These factors summarise information contained in multiple data series, thereby reducing model complexity. By capturing common movements across different series, the factor model simplifies the data structure without losing essential information and prevents overfitting. Moreover, this formulation of the model allows for the inclusion of missing observations, which is common in real-time datasets where not all variables are observed at every point in time.

The research by Modugno (2011) aims to explain the official index, which is updated with a monthly frequency, using data with a higher frequency. Therefore, the model parameters are estimated using a maximum likelihood method adapted to handle mixed frequency data sets. This method is robust to the presence of missing data and can be dynamically updated as new data become available.

In contrast, in our research, we compile our own weekly price index (CMPI), which is already high-frequency itself compared to the official indices. In explaining the CMPI through commodity prices, we do not have to adhere too strictly to that methodology (see Section 4.2). Its relevance for our research will only come into play when trying to estimate the official producer price index (see Section 4.3).

2.2.2 Official indices estimation

The second objective of our forecasting endeavor is devoted to explaining official producer price indices using our CMPI. Unlike the estimation of the CMPI itself, where we predominantly regress the weekly index on weekly data, the official indices are published monthly at best, creating a frequency mismatch that itself requires different methods.

Bertolotto & Aparicio (2020) use vector autoregression (VAR) models to analyse the dynamic response of CPI inflation to shocks in internet price indices. They perform forecasts at one, two and three month horizons using online price indices. These models outperform traditional benchmarks such as AR(1) model,[*)] random walk and Phillips curve models.

Macias et al. (2022) use the ARMA model adjusted for seasonality and augmented with exogenous variables (SARMAX) as a basic approach to forecast food inflation using internet price data. They use a recursive strategy to continuously update their forecasts as new data becomes available. This recursive approach is key to integrating the latest information and maintaining the accuracy of their real-time nowcasting model.

Powell et al. (2017) use a state space model that is inherently suitable for integrating data with mixed frequencies. This type of model allows them to model daily data while produ-

*) Autoregressive model, where the explained variable depends on the value immediately preceding the current value.

cing estimates that are consistent with the monthly frequency of CPI measurements. Their model conceptualizes two parallel processes of inflation – one based on traditional survey data and the other on data obtained from websites. Inflation estimates from both sources are modeled to interact, which increases the robustness of inflation forecasts. Furthermore, their work also introduces a sequential learning algorithm for updating inflation estimates.

Monteforte & Moretti (2009) use MIDAS (Mixed Data Sampling) regression models to integrate daily financial data with monthly core inflation indices. They do not use web-derived data for their estimates. Instead, their model includes both high-frequency financial variables (in our case, CMPI or commodity prices) and lower-frequency economic variables in the same regression framework. MIDAS models are equipped to handle the time mismatch between these types of data without having to aggregate the higher frequency data to the lower frequency data. They construct a monthly core inflation index using a generalized dynamic factor model (GDFM) that extracts common predictive factors from a large set of macroeconomic variables. This model captures medium to long-run inflation dynamics.

Macias et al. (2022) argue against MIDAS models despite their recent popularity because they often use nonlinear weighting schemes to integrate variables with mixed frequencies. While this method of incorporating daily data through polynomial lag structures is clearly effective for capturing dynamics in models that include high frequency variables, it is not consistent with the direct arithmetic averaging used in the official CPI calculation. Therefore, they first aggregate the daily data into a monthly index, which then enters the models.

2.3 Contribution

One of the main innovations of this thesis is the effective use of unstructured big data obtained from the web, which becomes a valuable input for economic forecasting. As described in the previous sections, this effort is not entirely new in itself, but put in context, our work extends existing knowledge in multiple ways.

To the best of our knowledge, compiling producer price indices using web-scraped data has not yet been performed, as all previous work has focused on CPI inflation (Cavallo & Rigobon, 2016; Breton et al., 2016; Powell et al., 2017; Bertolotto & Aparicio, 2020; Macias et al., 2022). Improving the measurement and predictability of producer prices naturally contributes to efficient monetary policy decision-making, but can also help the business sector when it comes to budgeting, e.g. for construction projects.

Similarly, no such research has yet been conducted in Czechia, even if we extend the scope from producer prices to consumer prices. Therefore, this thesis presents a unique study of the period covering the recovery from the covid-19 pandemic and the impact of the Russian invasion of Ukraine on the price level in this Central European open economy, when even the official indices were well off their long-term averages.

In this area, high-frequency data have so far been used to model official indices with lower frequency. This paper provides the reader with a first look at explaining the high--frequency index (CMPI) using high-frequency commodity price data. We examine the efficiency and usefulness of such forecasting efforts by putting the results in the context of new forecasts of official price indices.

We go beyond previous attempts to explain CPI inflation through commodity prices, which have provided rather inconclusive results in the case of the general inflation rate (Bosworth & Lawrence, 1982; Beckerman & Jenkinson, 1986; Durand & Blöndal, 1988; Boughton & Brenson, 1990), by linking it to PPI inflation.

But perhaps the biggest contribution of the thesis is the use of high-frequency web-scraped data to nowcast the official price index (MIPI), demonstrating that higher accuracy can be achieved as a result.

Chapter 3

Methodology

At its core, the research can be divided into two parts: data collection and the modeling of producer prices. This section adheres to this structure and describes the process of compiling price indices from web-scraped data and the econometric methods behind effective forecasts. The first part extends the insights gained during our previous research (Štefl, 2022) and takes the methods used to a more automated and standardized level. Subsequently, the second part builds upon the data assembled in the initial phase, leveraging it for further analysis and application.

3.1 Construction Materials Price Index

For the objectives of our research, we produced the Construction Materials Price Index (CMPI). This weekly index is derived from data systematically collected via web scraping from the websites of the three biggest online resellers of the construction materials in the Czech republic. Data collection at a reliable level started on 3 October 2021 after a preparatory period of about three months. By the conclusion of the project on 30 June 2024, this initiative yielded 144 observations spanning a period of 33 months.

3.1.1 Data processing

The data collection process begins with web scraping,[*)] which is essentially an automated method of extracting data from websites. There are several approaches, for the eshops we focused on, Python along with the *Beautiful Soup* package proved to be the most effective.

*) Principles of web scraping are explained in detail in our previous work (Štefl, 2022).

Every Sunday, the web-scraping robots automatically download the complete content of the three building material retailers' e-shops. The predefined data frame is the same for all eshops and includes the following fields in addition to system fields such as observation date or shop name: item name, item URL, parent groups (at three levels), item ID, price, discount if any, and availability. Although these data files are seemingly the same for every eshop, this is only partially true. The system of each website is different, so item names and item groups do not match and the final output of web scraping is not structured without

Figure 1: CMPI calculation system

October 2021 – June 2024

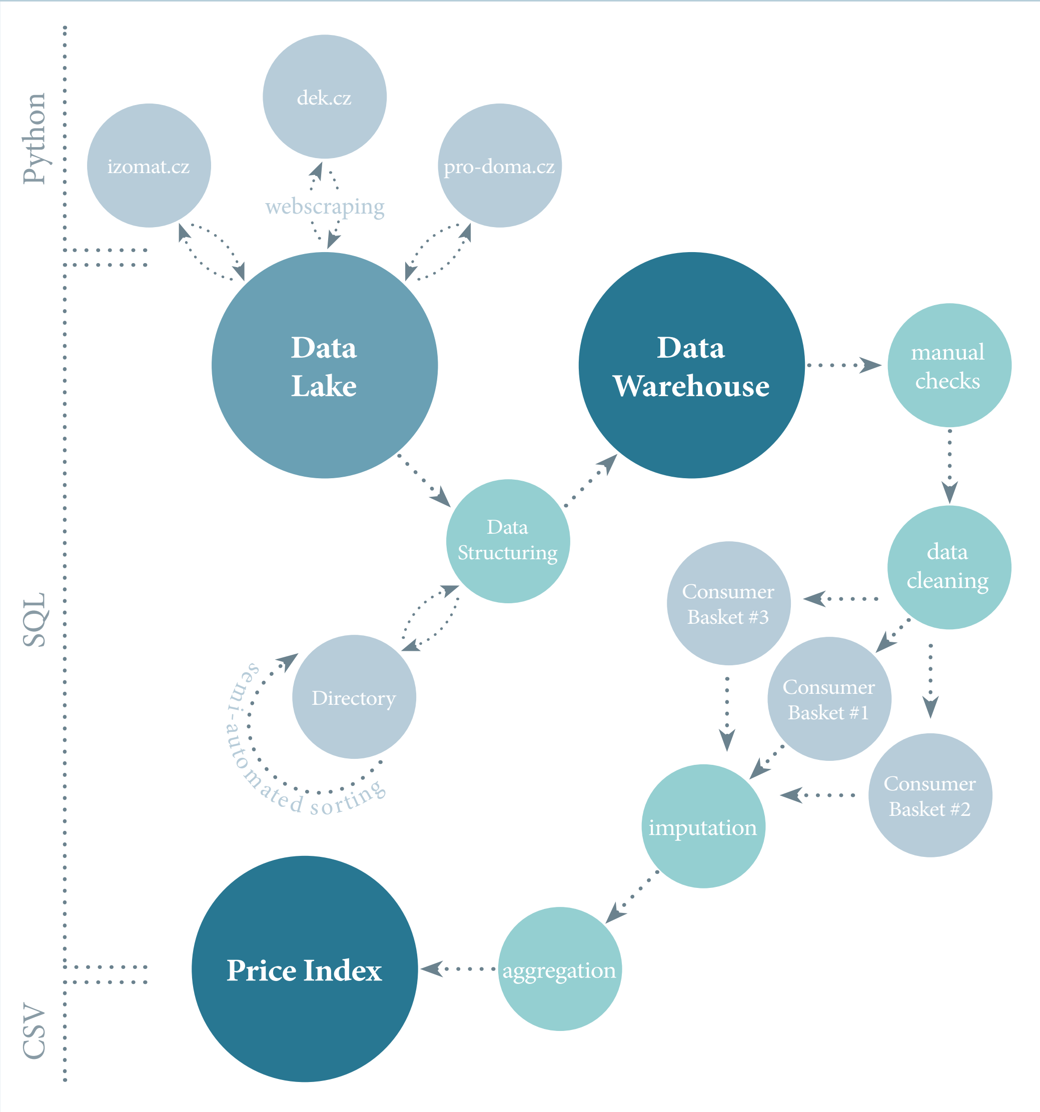


Figure 1 shows the CMPI derivation steps from data collection to the final index itself. The process takes place every week (except for the imputation, which is not necessary every time). Data collection and manual checks are the most time-consuming procedures, but the entire system is completed in less than four hours on average.

further data cleaning. But even the raw data stored in this data lake can be used for a lot of interesting research that is not necessarily directly related to inflation (Lünnemann & Wintr, 2006). Due to the immense amount of data collected in this way (see Table 1), downstream processes take place in SQL, where the main goal of the cleaning process is to distill a structured database which can easily be used for our research.

In the semi-automated sorting process, each item is first tagged with a CMPI ID that assigns the item a place in a unified four-level item group system, regardless of which eshop it was downloaded from (Štefl, 2022). After a series of validation mechanisms in this way, a structured panel database of all items from all eshops over the entire research period is created.

CMPI database

The data is stored in a relational database because the amount of data collected would make it impractical to process it in any other way. Over the 11 quarters covered by the project, more than 12 million unique records were collected.

Table 1: Summary of CMPI database content
October 2021 – June 2024

Eshop	# of products	# of observed prices	# of prices per week (AVG)
Dek	115,371	9,425,245	65,453
Pro-Doma	16,194	1,464,578	10,241
Izomat	20,101	1,832,487	12,725
Total	**151,666**	**12,722,310**	**88,419**

Table 1 provides a comprehensive summary of the data web-scraped from the websites we focusesd on from October 2021 to June 2024. The figures refer to the cleaned database and therefore only show distinct records.

The CMPI database contains the full offer of construction materials listed on the selected eshops. However, this is not the entire offer of these eshops, as some of them contain a wider range of products, such as smart home appliances, workwear or cleaning products. These items, which are clearly outside the scope of our research, were avoided as far as possible in our data collection as they would only artificially increase the size of the database.

Although automated cleansing processes aim to keep the data stored in the database consistent, they may not be robust enough for such large data. Therefore, all important fields such as item classification or packaging size are also checked manually for items selected for the consumer basket.

Given the amount of data in the database, aggregate information on price dynamics are only a guide to movements within the consumption basket, which is relevant for our research in the first place. The average time we observe an item is 92 weeks in the adjusted data (i.e., not including outliers such as items that only appeared for one week, etc.) and on these items the price remains unchanged for an average of 30 weeks (see Table 2).

Table 2: Price dynamics of the CMPI database

October 2021 – June 2024

Eshop	mean # of weeks a price is observed	median # of weeks a price is observed	mean # of weeks a price is unchanged	median # of weeks a price is unchanged
Dek	91	80	31	24
Pro-Doma	91	101	24	19
Izomat	93	109	29	26
Average	**92**	**87**	**30**	**24**

Table 2 summarizes the price dynamics of the cleaned data accross the three eshops. The comparison shows that the shortest item appearance time is found at Dek, which also offers on average more than five times more items than the other eshops. This is a summary of the entire database, the values are different for the consumption basket, which is the primary basis of our research.

3.1.2 CMPI compilation

We refer to the core index of our research as the Construction Materials Price Index (CMPI). For its calculation, we closely follow the official methodology established by the CZSO, with a few justified deviations, which are described in the following lines. The CMPI aims to be a proxy for the official *Price Index of Material Inputs to Construction Work* (MIPI) published monthly by the CZSO. In addition to this standard producer price index, we compare the official index with several alternative unweighted indices, for example using the Jevons approach.

First, 600 items are selected from the cleaned data into a predefined basket of goods for each eshop separately (i.e., 1,800 items for three eshops). Following the methodology of Štefl (2022), the weights of the items are set to create a basket of building material goods that would be needed for the construction of a typical single-family house. This selection does not exactly match the goods selected by the CZSO which may cause a possible *product selection bias*. Since the official methodology collects prices via surveys by contracted construction works, given the different primary data source, we do not try to circumvent this potential disparity in the input, but account for it in the following analyses to minimize its impact. Furthermore, before the CMPI itself is produced, the resulting dataset of 1,800 items is partially manually checked to ensure consistency of data points across all periods. This check proved essential during the research to avoid *product classification error*. Despite the advances, we have not been able to replace this manual check with a fully automated process and, together with Macias et al. (2022), we consider this area as a suitable application of machine learning.

These three indices based on the same consumer basket are then combined into a final CMPI using the arithmetic mean. This approach allows us to see the data at a further level of detail so that we can compare price trends across individual stores. On the other hand, this is an extra step compared to the official methodology, which works with monthly questionnaires sent to construction companies. However, the CZSO also averages these individual surveys, so we consider our approach to be similarly representative, as it tries to cover as large an area of the market as possible.

Index calculation

After processing the structured data, the index calculation is fairly straightforward. In line with the CZSO methodology, we use a Laspeyres approach to obtain the CMPI, i.e., we keep the consumption basket fixed and revalue it in each period as described above. The resulting index formula can therefore be expressed as

$$I_t = \frac{\sum_{i=1}^{n}(p_{it} \cdot q_{i0})}{\sum_{i=1}^{n}(p_{i0} \cdot q_{i0})} \cdot 100 \ , \tag{3.1}$$

where q_{i0} denotes the quantity of an item i in the base period, p_{i0} its initial price and p_n its price at the n-th period (Mankiw, 1997). We report the individual quantities and their associated constant base period weights $p_{i0}q_{i0}$ in Appendix C.

Missing items

The trade-off between data quantity and data quality obviously shifts to data quantity for web-scraped data by definition as we can instantly collect huge amounts of information, but at the quality specified by the respective website. In addition to sorting into the correct categories, the cleaning and sorting processes are also supposed to resolve allegedly missing records, which occur when an e-shop changes, for example, the name or internal code of an item, and it then falsly appears to be completely new. The importance of proper data processing mechanisms is thus evidenced in our research by the fact that the proportion of missing price quotes drops from 24% to 7% after cleaning.

Although the intensive sorting processes create a structured dataset, they cannot resolve all missing records. The remaining 7% is either due to the failure of web-scraping robots, or simply because the item is actually missing from the e-shop's offer at that point of time. To ensure the consistency of the index calculation, these two cases need to be handled by different methods.

In the first mentioned case, if there is an error in scraping from the web, according to the methodology presented by Štefl (2022), a delay of up to three days is considered negligible, because prices usually change every 30 weeks (see Table 2), so such a delay is negligible. If the problem persists for more than three days, the website is scraped again in the next scheduled week. In nearly three years of research, all software problems that have occurred have been minor and have been resolved within the three-day limit, so a contingency plan for longer delays has never really been needed.

If an item is missing because it was actually unavailable at the time, we follow the official methodology and switch to imputed prices. The ideal solution is to find a substitute product that matches the original in key aspects such as size, price and performance, and differs only in less important features such as colour or manufacturer. This helps maintain consistency in data quality. If no suitable substitute can be found, a ratio imputation method is used using the geometric mean of the growth, which can be expressed as

$$\hat{p}_i^t = p_i^{t-1} \cdot \left(\prod_{\substack{j \in I \\ j \neq i}} \frac{p_j^t}{p_j^{t-1}} \right)^{\frac{1}{n-1}} \tag{3.2}$$

This method calculates growth rates based on comparable items within the same product category to minimize bias in the price index. This approach continues until either the original item is available again or an appropriate replacement is found, with a preference for actual data over imputed values whenever possible.

Alternative indices

In addition to the CMPI, which follows the official CZSO methodology as far as it is possible with web-scraped data, we compile two alternative indices that are not primarily based on the consumer basket. We outlined the procedure for constructing them in the appendix of a previous work (Štefl, 2022), where in addition to a naive but easily producible indicator, which we called the *unsorted mean index,* we also introduced the *unit price* and *chained weekly index*. Unlike our core CMPI, the methodology of which was introduced in the previous lines and is based on the Laspeyres approach, these two are based on the Jevons methodology. Like Breton (2022), we will refer to the *unit price index* as *fixed-based Jevons* (CMPI_{fix}) and the *chained weekly index* as *chained bilateral Jevons* ($\text{CMPI}_{\text{chain}}$) for the sake of clarity.

Both indices neglect the quantity of items consumed and therefore abstract from the weighting system. While this is not at all representative for standard data collection, for the web-scraped data, where we are able to collect the entire offer of an e-shop – in theory the entire market – such indices answer the question of what is the expected inflation of a randomly selected good. As their names suggest, both are based on the Jevons approach, which can be generalized as a geometric mean of price changes of individual items

$$P_J = \left(\prod \frac{p_t}{p_0}\right)^{\frac{1}{n}}. \tag{3.3}$$

The *chained bilateral Jevons index* compares the weekly peers for all items from the cleaned data (i.e., all outliers are omitted) and uses all the feasible web-scraped data. On the contrary, the *fixed-based Jevons index* limits the number of items as it only counts items that occurred during the entire research period.

Despite their very different meaning compared to the official indices, both $\text{CMPI}_{\text{chain}}$ and CMPI_{fix} represent a viable option to produce a price index free of any imputation bias or product classification errors and also do not suffer from weight changes in the consumption basket.

Evaluation of indices

A comparison of the CMPI, the unweighted indices and the official MIPI is shown in Figure 2. Similar to Macias et al. (2022), the unweighted indices also provide only a weak approximation of the official index in the case of producer prices, as they both significantly understate the MIPI.

For most of the time, the CMPI performs very well under the MIPI overlay, with the exception of the period at the turn of 2022 and 2023. We attribute this one-off difference, which affects the overall inflation rate for building materials, especially when expressed in base terms, to the change in the weighting system in the CZSO methodology from January 2023.

In addition to the Pearson correlation coefficient, to quantify the similarity, and therefore to assess the usefulness of the index, we use the root mean square error (RMSE), which is defined as

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T}}. \tag{3.4}$$

We also use the same metric to evaluate the performance of our nowcasting exercise in the following sections. For our purposes, it is particularly useful because it gives relatively large weight to large errors due to the square of each term, which means that large errors have a disproportionately large effect on the RMSE. Thus, the RMSE is sensitive to outliers.

A comparison of these three indices with the official MIPI is shown in Figure 2. Contrary to expectations, unweighted indices performed well in addition to the standard CMPI and, according to the above criteria, the averaged monthly $\text{CMPI}_{\text{chain}}$ even outperformed the CMPI in fitting the MIPI without considering lags. The proximity of the unweighted Jevons indices to the Laspeyres indices with a fixed-base consumption basket represents a

Figure 2: Performance of alternative CMPI indices compared to MIPI

October 2021 – June 2024, month-on-month index, MIPI reported at mid-month, percent

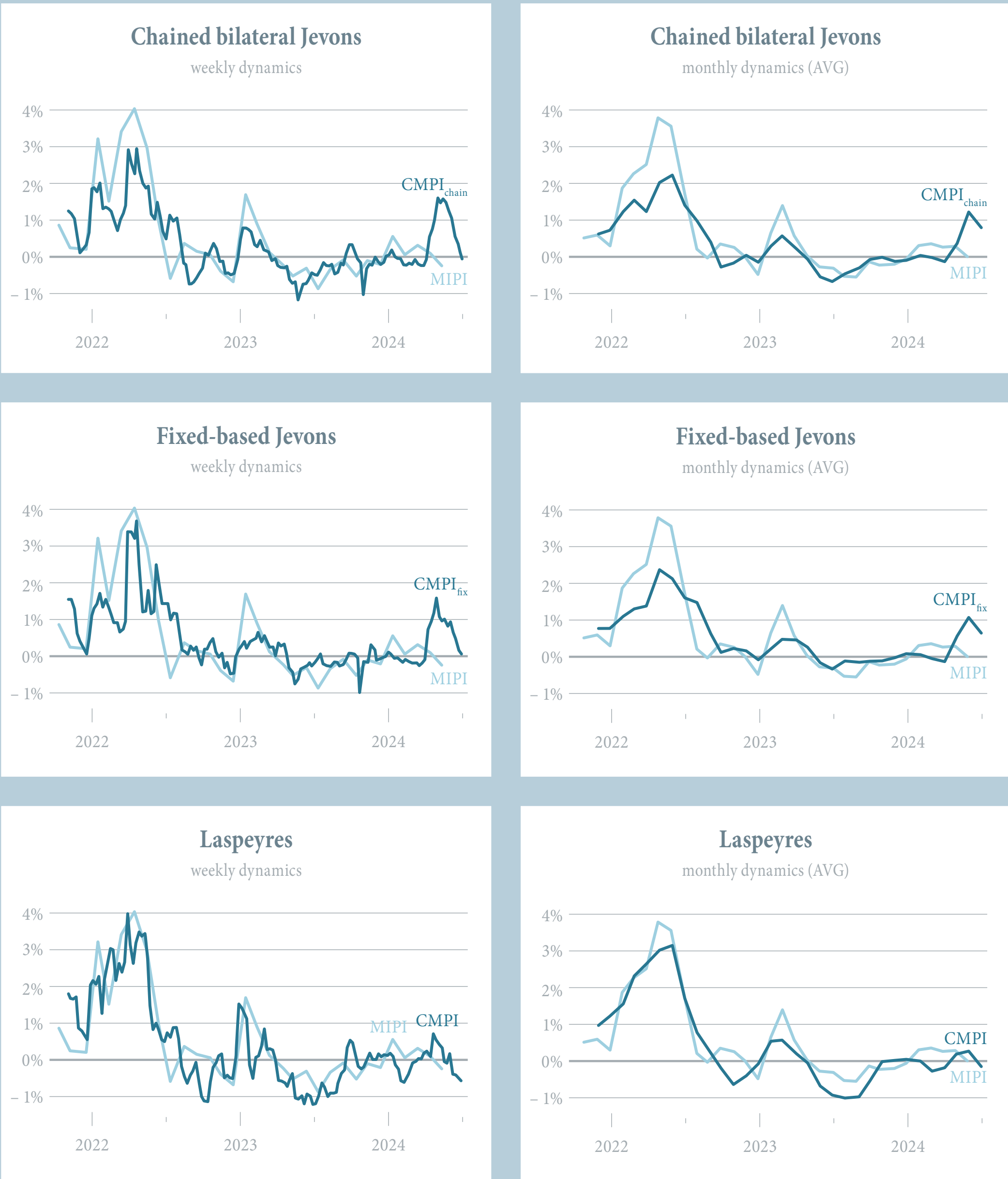


Figure 2 shows a comparison of the official MIPI and the three versions of the CMPI (all as month-on-month indices). The first two rows compare the unweighted Jevons indices ($\text{CMPI}_{\text{chain}}$, CMPI_{fix}) and the last row studies the CMPI compiled according to the Laspeyres approach. The left column tracks weekly dynamics, while the right column tracks monthly dynamics of the averaged indices.

different finding than that reported by Macias et al. (2022) for the CPI. This suggests that for less complex price indices, such as producer prices index components, the unweighted framework may work reasonably well. Given the considerably simpler complexity of compiling such an index, which does not involve cleaning and theoretically may not need the sorting process, these types of indices are worth pursuing in any follow-up research for their overall evaluation.

Exploring the relationship between the official MIPI and our CMPI begins with the question of proper temporal overlap. Assuming that the indices are published two weeks after the end of the reference period, we can conclude that the CMPI level obtained in the week of publication is exactly one month ahead of the official monthly index. This finding is supported by the highest value of the correlation coefficient and the lowest RMSE for overlapping MIPI and CMPI with a one-month lag.

Table 3: CMPI fitting performance of official index

October 2021 – June 2024, month-on-month, monthly average, percentage points (RMSE)

Index framework	correlation with MIPI	RMSE
Laspeyres (CMPI)	0.93	0.43
fixed-based Jevons (CMPI_{fix})	0.85	0.62
chained bilateral Jevons ($\text{CMPI}_{\text{chain}}$)	0.84	0.63

Table 3 shows the correlation of each type of CMPI with the official MIPI and the associated root mean square error (RMSE). Contrary to expectation, both Jevons-type indices show a surprisingly good ability to fit the official index. However, the Laspeyres approach outperforms both, confirming the role of CMPI as a pivotal indicator for our nowcasting research.

With this assumption, by studying month-on-month inflation, the simple correlation between the MIPI and the worst-fit Jevons-type index $\text{CMPI}_{\text{chain}}$ reaches 0.84, and the RMSE is equal to 0.63 p.p. in this case. For comparison, the best-fit online index in Macias et al. (2022) showed the same correlation (0.77) and an RMSE of 0.50 p.p. However, we should also add that their research covered a time period three times longer than ours, so the comparison is only indicative.

Although the differences between the Jevons and Laspeyres index types are very small in the overall context, the Laspeyres-type CMPI still performed best in matching the MIPI, thus confirming its role as a key index for our research. The joint correlation amounts to 0.93 and the RMSE equals 0.43 p.p. in this case.

3.2 Data for explanatory variables

At a basic level, the data we use for our research can be divided into three data sets. In addition to the CMPI, the compilation of which is described in Section 3.1, and the official CZSO indices, we selected 11 energy or material commodities to study the interdependence of producer prices.

By their very nature, most of these variables are strongly interconnected, since, for example, petrol and diesel are derived from oil, as well as oil and natural gas are imported to the Czech Republic mainly from the same countries and their price has skyrocketed after the Russian invasion of Ukraine in 2022, which is also true for electricity, etc. Data on material commodities were obtained through the Refinitive Eikon Datastream application,[*)] and electricity prices related to the Czech market were extracted from Power Exchange Central Europe.[**)]

Having obtained a set of daily price quotes, we calculated the intraday differences, which were then aggregated and properly linked to the CMPI. Given that CMPI data collection began in October 2021, we study the dataset of explanatory variables from January 2021 onwards to allow for lagged effects analysis.

We also considered the datasets used by Modugno (2011), which included high-frequency data sources such as the *Weekly Commission Oil Bulletin Price Statistics*, *Weekly Retail Gasoline and Diesel Prices*. We ultimately chose to rely solely on price quotes from commodity exchanges, i.e., on primary data, which is more in line with the actual nature of the web-scraped data.

The list of these variables includes: electricity, crude oil (Brent), crude oil (WTI), petrol, diesel, natural gas, aluminium, copper, iron ore, lumber, steel. In addition to data availability and frequency, expected dependencies played a role in the initial selection of commodities.

The cost of electricity directly affects the cost of domestic production of all building materials, which should impact the overall price level. Brent and WTI crude oil are both main indicators of world oil prices, which affect the cost of fuel required for transport and machinery used in the production of building materials. However, in addition to its use in transport, it is also a primary raw material for the production of plastic materials that can be imported into the Czech Republic, which is why we include the price of WTI crude oil, although Brent crude oil should be more relevant for the Czech Republic in other aspects.

[*)] https://www.refinitiv.com/en/products/datastream-macroeconomic-analysis

[**)] https://pxe.cz/cs/derivatovy-trh/elektrina

Petrol and diesel are crucial for transport logistics, not only in the building materials supply chain, and affect the total cost of delivery. In this case we use prices relevant for the Czech Republic. Natural gas is often used as a primary energy source in manufacturing processes, including the production of materials such as glass and bricks. Moreover, like other energy commodities, its prices have experienced turbulent developments in the period under review following the Russian invasion of Ukraine. Aluminium, copper, iron ore, steel are all metals used as basic raw materials directly in the construction industry. The same is true for lumber, where the price of wood as a basic building material affects a significant proportion of building materials, especially roofs and floors.

Table 4: Summary statistics of week-on-week changes of variables

October 2021 – June 2024, week-on-week, percent

Variable	Mean	SD	Median	Min	Max	Skewness	Kurtosis
CMPI	0.09	0.45	$8.6 \cdot 10^{-6}$	−0.75	1.81	1.97	8.15
Aluminum	−0.06	2.95	−0.17	−7.26	10.57	0.44	4.04
Copper	0.06	2.58	0.21	−6.39	8.05	0.30	4.06
Crude Oil Brent	0.15	4.01	0.30	−9.93	15.17	0.18	4.45
Crude Oil WTI	0.17	4.31	0.42	−10.63	18.78	0.31	4.91
Diesel	0.13	2.49	−0.15	−4.55	19.14	3.75	27.25
Electricity	0.36	9.74	−0.59	−41.69	51.32	0.62	10.02
Iron Ore	0.00	3.92	0.13	−20.03	10.50	−0.96	7.45
Lumber	0.02	7.09	−0.97	−22.96	29.03	0.82	5.93
Natural Gas	0.41	9.62	−0.47	−40.94	29.03	−0.03	6.43
Petrol	0.10	1.98	−0.02	−5.04	14.10	2.84	20.69
Steel	−0.45	5.35	−0.92	−11.94	30.35	1.77	10.76

Table 4 presents a comprehensive statistical analysis of various commodities and their comparison with the CMPI. The very low median of the CMPI, contrasted with its standard deviation and the range of minimum and maximum values, suggests a distribution where price changes are usually small but can sometimes be significant. Notably, *Diesel* and *Petrol* show significantly high kurtosis (27.25 and 20.69), indicating considerable occurrence of outliers.

In addition to these commodity prices, we also decided to examine the impact of the CNB repo rate, which we obtained directly from its website.*) However, due to the high frequency of CMPI data and commodity prices, it has not proven to have a positive impact on the goodness of fit of the models (see Section 4.2).

*) https://www.cnb.cz/en/faq/How-has-the-CNB-two-week-repo-rate-changed-over-time

3.3 CMPI estimation

This section summarizes the theory of econometric models used in our research. In addition to the autoregressive integrated moving average (ARIMA) model that we use for the estimation of our CMPI by itself, we also present here its extension with exogenous variables (ARIMAX) and the vector autoregressive (VAR) models. We use the latter two models to estimate the CMPI through the additional commodity variables introduced in the previous section. Besides introducing the individual models, we also add a description of a framework for assessing the implications of the different variables and the model evaluation criteria.

3.3.1 ARIMA model

In an effort to estimate the CMPI, ARIMA models represent the cornerstone of our endeavor. This set of statistical models is primarily used for single, univariate time series data that show signs of non-stationarity. It contains three key components: autoregressive (AR), differential (I – for integrated) and moving average (MA). In essence, ARIMA can be considered the most general category of models explaining time series that are or can be converted to stationary.

For completeness, let us define the corresponding argument for each of the listed components:

- p – the number of autoregressive terms (lag degree),
- d – the number of nonseasonal differences required for achieving stationarity, and
- q – the number of lagged prediction errors.

The autoregressive component explains the selected variable by its own lagged values. Its simplest version for $p = 1$ is the ARIMA(1, 0, 0). This is also known as the AR(1) model, which covers a one-period delay, i.e., in this case the previous observation explains the following, according to the equation

$$Y_t = c + \phi Y_{t-1} + \epsilon_t, \tag{3.5}$$

where c is a constant, ϕ stands for the slope coefficient quantifying the influence of the lagged value Y_{t-1} and ϵ_t is the error term.

The general formula for the autoregressive process of a p-th degree can be defined as

$$Y_t = c + \textstyle\sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t, \tag{3.6}$$

where Y_t is the explained variable, c is a constant, ϕ_i stands for the slope coefficient of i-th lagged Y_t and ϵ_t is the error term.

The differencing component proves its strengths when the model has to deal with non--stationary data. A time series is considered stationary if its statistical properties, namely the mean and variance, do not change over time. This means that a stationary time series does not follow a trend, fluctuates with a consistent magnitude around its mean and shows reliable cyclical patterns over time. The simplest differencing model suitable for non-stationary data with $d = 1$ is ARIMA(0, 1, 0), which essentially is a random walk model with a prediction equation expressed as

$$Y_t = c + Y_{t-1} + \epsilon_t, \tag{3.7}$$

with all symbols keeping the same meaning as in previous two equations.

In this context, a time series can generally be thought of as a mixture of signal and noise. Signal can represent underlying patterns such as trends, oscillations or seasonal effects, while noise is by definition random and unpredictable, as captured in the equation 3.7. The ARIMA model essentially acts as a filter that separates the signal from the noise and uses the filtered signal to predict future values.

To this end, non-stationary data can be converted to stationary data by differencing, the first stage of which is essentially a subtraction of adjacent values according to the formula

$$\Delta Y_t = Y_t - Y_{t-1}\,. \tag{3.8}$$

For the second stage of differencing we obtain the formula

$$\Delta^2 Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}\,. \tag{3.9}$$

This process can continue and differencing operator Δ can be generalized for higher degrees in the backward shift operator B as

$$B^k(Y_t) = Y_{t-k}\,, \tag{3.10}$$

where k is the number of periods the time-series is shifted backward.

The moving average component, represented by the parameter q, involves expressing the time series as a function of its past error terms, which essentially captures the persistent effects of past shocks that were not captured by the autoregressive part of the model. The general formula for its q-th degree can be written as

$$Y_t = c + \textstyle\sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t\,, \tag{3.11}$$

where Y_t is the explained variable, c is a constant (or the mean of the series), θ_i stands for the slope coefficient of i-th MA term, indicating the impact of the error terms from i periods ago on the current value and ϵ_t is the error term at time t, representing white noise with a mean of zero and a constant variance.

ARIMA models are widely used in time series forecasting due to their flexibility and robustness, but like any model they have their strengths and weaknesses. Their greatest strength is perhaps their versatility and comprehensiveness, as they can handle a wide range of time series, whether seasonal or non-seasonal, and can capture both the moment and mean reversion characteristics of the data, making them effective for both short-term and long-term estimation.

One of the main limitations of ARIMA models is the need for the data to be stationary. This often requires transformations such as differencing (which can be handled directly through the model settings), logarithming or detrending, which can complicate the model building process and interpretation of results. Thus, their introduction must be preceded by statistical testing of the underlying data, which we describe in the following sections.

A fundamental flaw for some cases is that ARIMA models are based primarily on the assumption that future patterns will replicate past patterns. This may limit their effectiveness in situations where the time series is affected by new factors or structural changes that are not accounted for by historical data. However, extended versions such as ARIMAX can address this by augmenting the model with exogenous variables.

3.3.2 ARIMAX model

The ARIMAX (Autoregressive Integrated Moving Average with eXogenous variables) model is an extension of the traditional ARIMA model that includes external or independent variables in the forecasting equation. This modification makes it possible to address some of the limitations of the standard ARIMA model, in particular its inability to incorporate influences outside the forecast data set. Since the CMPI data contain their week-on-week representation, the time series is already *naturally* differenced, and adding selected commodities as exogenous variables is sufficient to increase the forecast accuracy of the model.

In general form, the formula of the ARIMAX model can be formulated using equations 3.6 and 3.11 as follows

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{k=1}^{r} \beta_k X_{k,t} + \epsilon_t, \tag{3.12}$$

where besides the already introduced AR and MA components, we add a component for r external variables, with coefficients β. The equation can be altered using the equation 3.10 for backward shift operator as

$$(1 - B)^d Y_t = c + \left(\sum_{i=1}^{p} \phi_i B^i\right) Y_t' + \left(\sum_{j=1}^{q} \theta_j B^j\right) \epsilon_t + \sum_{k=1}^{r} \beta_k X_{k,t} + \epsilon_t, \tag{3.13}$$

where $(1 - B)^d Y_t$ represents the d-th difference of the series Y_t.

Although the inclusion of external variables can solve the aforementioned weakness of ARIMA models, which occurs when the regressand cannot be explained by itself, it is associated with three additional challenges.

Clearly, the complexity of model selection and calibration increases. Determining which external factors to include and determining their lagged effects can be demanding and requires a deep understanding of the model and the domain.

As the number of estimated parameters rises, the risk of overfitting increases, especially if the sample size is not large enough or the selected exogenous variables are not truly predictive of movements in the dependent variable.

Moreover, the efficiency of the ARIMAX model depends largely on the availability and quality of data for both the dependent series and the exogenous variables. Any inconsistencies in the data or missing values can significantly degrade the performance of the model. While the latter is not a problem for our model because we use cleaning and imputation methods to prevent data inconsistency (see Section 3.1.2), the first two issues need to be overcome when fitting the model.

3.3.3 Dynamic factor model

One way to reduce the dimensionality of our model with commodity prices is to use a *dynamic factor model* (DFM). The connection between individual commodity time series and inflation itself is complex and assessing their importance for the ARIMAX model is not always straightforward.

Dynamic factor models are designed to explain differences between multiple interrelated time series through a few unobserved common factors and possibly some idiosyncratic terms. These models assume that many time series can be controlled by a smaller number of hidden factors that capture the common variability of the data. This makes it easier to analyze and interpret the underlying dynamics, but at the cost of computational intenstity. Dynamic factor models are usually more complex and computationally demanding because unobserved factors need to be distilled from large datasets and estimated individually.

The basic interpretation of the dynamic factor model can be written as

$$Y_t = \Lambda F_t + \epsilon_t \,, \tag{3.14}$$

where Y_t is an $n \times 1$ vector of observed time series at time t, Λ is an $n \times r$ matrix of factor loadings, which maps each factor onto the observed time series, F_t is an $r \times 1$ vector of common latent factors at time and ϵ_t is an $n \times 1$ vector of idiosyncratic errors for each time series at time t.

The vector of unobserved components F_t can also include the autoregressive dynamics, which can be expressed as

$$F_t = \Phi F_{t-1} + \eta_t, \tag{3.15}$$

where Φ represents the autoregressive coefficient matrix and η_t is the vector of shocks or factor changes at time *t*, which are usually assumed to be white noise.

3.3.4 Granger causality test

Assuming that we include more than ten variables in the models, it is crucial to determine whether the selected exogenous variables actually provide predictive information for the dependent time series beyond what is already available from the series' own past values. To do this, we use the *Granger causality test*, which is a statistical hypothesis test based on the principle that if variable *X* provides statistically significant information that helps predict future values of the variable *Y*, then *X* is said to *Granger-cause Y*.

It is important to note that Granger causality does not imply true causality in the sense of real dependencies. Instead, it suggests predictability, i.e., if knowing the values of *X* helps predict *Y* better than merely knowing past values of *Y*. There may always be another *hidden* relationship that affects both *X* and *Y*, or it may involve a spurious correlation.

The process of testing involves estimating two models. The restricted version explains the regressand *Y* only on its own lagged values as

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \epsilon_t, \tag{3.16}$$

while the unrestricted model includes also the past values of *X*:

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{i=1}^{q} \gamma_i X_{t-i} + \epsilon_t. \tag{3.17}$$

The null hypothesis states that if the coefficients γ_i of the lagged values of *X* are zero for all *i*, *X* does not Granger-cause *Y*. But if at least one of these coefficients is non-zero, the alternative hypothesis implying Granger causality between *X* and *Y* holds.

The Granger causality test is crucial in our research for both the ARIMAX model and the dynamic factor model. In both cases, it assesses the importance of exogenous variables. However, in the case of the ARIMAX model, it determines whether the potential exogenous variables actually provide predictive information for the dependent time series beyond that already available from the series' own past values, i.e., whether it would not be sufficient to stay with the ARIMA model alone.

3.3.5 Criteria for model selection

To decide which model is better for our purposes, we use the *Akaike information criterion* (AIC) and the *Bayesian information criterion* (BIC). These criteria are essential for comparing the performance of the models and help to balance their fit and complexity to avoid overfitting and underfitting, thereby providing more reliable predictive power.

Akaike information criterion

This criterion is particularly useful for the selection of the fitted model. It provides comparative information on model quality by balancing measures of goodness of fit and model complexity (in terms of number of parameters).

It is calculated using the formula

$$\mathrm{AIC} = 2k - 2\ln(L), \tag{3.18}$$

where k is the number of model parameters and L is the model likelihood which measures how well the model fits the data.

In the context of ARIMAX models, AIC can help determine the optimal number of lags and whether including certain external variables improves the model. To avoid overfitting, formula (3.18) penalizes the number of parameters by an expression of $2k$ while subtracting twice the log-likelihood, which becomes a minimization problem because a lower AIC value generally implies a better model as long as the difference is significant.

For the small-sized samples, the AIC can be converted in to corrected Akaike information (AICc) by adding the correction factor accounting for the sample size n

$$\mathrm{AICc} = \mathrm{AIC} + \tfrac{2k(k+1)}{n-k-1}\,. \tag{3.19}$$

The correction factor becomes particularly important when the sample size n approaches the number of parameters k. Once n is substantially larger than k, the correction term approaches zero and AICc converges to AIC.

It is generally recommended to use AICc over AIC when the sample size is small. While this is not the case for our ARIMA model, if the ratio of n over k falls below 40, we use AICc instead of AIC to evaluate the models. At the weekly granularity of our data, this is true from four or more parameters. For monthly data, this is true for all models (including simple ARIMA).

Bayesian information criterion

Similar to the Akaike information criterion, the BIC helps balance model complexity and its fit to the data, but applies a stronger penalty for models with more parameters, so it often favors simpler models over the AIC, especially as sample size increases.

The BIC formula is essentially similar to the AIC formula, but the number of data points is log-transformed, causing the penalty for model complexity to increase with the number of parameters and the logarithm of the sample size. The resulting formula is then

$$\mathrm{BIC} = k \cdot \ln(n) - 2 \cdot \ln(L), \tag{3.20}$$

where k stands for the number of parameters in the model, $\ln(n)$ is the natural logarithm of the number of data points (sample size), and $\ln(L)$ is the natural logarithm of the likelihood of the model given the data. As with AIC, when comparing models, the model with the lowest BIC is generally preferred.

Diebold-Mariano test

While both AIC and BIC are used to evaluate the fit and complexity of a single model, the *Diebold-Mariano* (DM) test directly compares the performance of two models based on their prediction errors.

The test is carried out by analysing the differences between the prediction errors of the two models. If one of the models consistently provides smaller prediction errors than the other, the DM test is likely to reject the null hypothesis, indicating a significant difference in prediction accuracy.

For each time period in the data set, the forecast errors for the two models are first calculated, subtracted from each other, and then a test (t-test in our case) is performed on the average of these differences. Under the null hypothesis, which states that both models have the same prediction accuracy, the expected value of the average difference is zero. The rejection of the null hypothesis implies that there is a statistically significant difference in the prediction accuracy of the two compared models. The sign of the DM test statistic indicates which model has better predictive performance. A positive DM statistic indicates that the first model is more accurate and vice versa.

Unlike the AIC or BIC, the DM test does not directly account for overfitting. The model may show better performance on the historical data used in the test, but this may not be true in general, and with additional observations the model may actually perform comparatively worse.

3.3.6 Lasso regression

Another method that helps us decide between possible models is the least absolute shrinkage and selection operator (Lasso) regression. Although Lasso is particularly renowned for its efficiency in high-dimensional settings (i.e., where the number of variables exceeds the number of observations, which is not our case), its advantages also apply to more traditional data configurations.

It contains a penalty term which is essentialy the absolute value of the magnitude of the coefficients. This feature helps in variable selection by reducing (shrinking) the less important coefficients to zero, effectively selecting a simpler and more interpretable model that avoids overfitting. Moreover, if multicollinearity is present (some regressors are correlated), Lasso can help by selecting one or more variables from the group of correlated variables and reducing the others to zero. This helps to reduce the variance of the model without substantially increasing the bias.

The Lasso regression minimizes the following function

$$\min_\beta \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}, \tag{3.21}$$

where β represents the vector of coefficients of the predictor variables, y_i stands for the observed outcome of the i-th observation, x_{ij} is the value of j-th predictor variable and λ is the regularization parameter. A higher λ increases the penalty and may lead to more coefficients being set to zero, thus simplifying the model.

3.3.7 Accuracy evaluation

To quantify errors, we use standard metrics, namely mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE), which we have already introduced in Section 3.1.2. Unsurprisingly, for all of them, the lower the value, the better the estimate fit.

In essence, the MAE is the average of the absolute differences between the prediction and the actual observation in the test sample, where all individual differences have equal weight. It is calculated using the formula

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|. \tag{3.22}$$

In the case of MSE, the errors are quadratic compared to MAE as given by

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{3.23}$$

The RMSE is then defined as the square root of the MSE as specified in equation 3.4 (see Section 3.1.2) and, as with the MSE, gives a relatively high weight for large errors. However, unlike MSE, it retains the original units, which makes it easier to interpret.

It is worth mentioning that MSE and RMSE are more sensitive to outliers than MAE. Hence, in datasets such as ours, where outliers are unlikely to be errors, these measures may be more appropriate than MAE. On the other hand, MAE is in the same units as the data, as is RMSE, while MSE is in square units of the data.

3.3.8 Cross-validation of predictive performance

Another method used for the overall evaluation of the model involves testing its predictive power in a pseudo real-time experiment. We use expanding and rolling window forecasts for this purpose.

Expanding window forecast

In cross-validation with an expanding window, the training window starts with an initial size and grows larger with each iteration, always including all previous data points up to the current time point. This method uses the entire data history as the window expands. The model continuously learns from the entire data history, allowing it to adapt to any structural changes or trends that have occurred over time. The influence of outlying or anomalous data points may diminish as the size of the training set increases, reducing their impact on model performance.

Rolling window forecast

In rolling-window cross-validation, a fixed-size window of data is used to train the model, and this window moves forward over time. This approach ensures that only the most recent data is used, which may be more relevant for predicting future values. Rolling window forecasts adapt to current changes in the data, making them particularly useful in environments where current information is more relevant than older data. This is advantageous in volatile markets or generally when modelling systems that change rapidly.

3.3.9 Stationarity testing

Both the ARIMA and DFM models as well as most time series models assume that the statistical properties of the data do not change over time – stationarity. This assumption allows the prediction formula to be defined in such a way that it can rely on the premise that patterns from the past will continue into the future.

We distinguish two types of stationarity according to their intensity. A time series is considered *strictly stationary* if its joint probabilistic properties are invariant to time shifts, i.e., the distribution of $(X_{t1}, X_{t2}, \ldots, X_{tk})$ is the identical to $(X_{t1+h}, X_{t2+h}, \ldots, X_{tk+h})$ for any integers $t1, t2, \ldots, tk$ and any shift h. In other words, if we take any segment of a strictly stationary time series and shift it along the time axis, it should look statistically indistinguishable from the original segment. This is a considerably strong condition that can rarely be achieved in real data.

A time series is *weakly stationary* if its mean and variance are constant over time and the covariance between two time points in the series depends only on the distance or lag between these points, not on the actual time at which these points are observed (sometimes referred to as second-order stationarity). This is more commonly used and most tests and analyses refer to it when talking about stationarity. This is more commonly used, and most tests and analyses refer to it when talking about stationarity, including the *Augmented Dickey-Fuller* (ADF) test that we use.

The ADF test examines the null hypothesis that a unit root is present in the time series sample. A unit root means that the series can be represented as a random walk or that it has some time-dependent structure. A significantly lower test statistic than the critical value leads to the rejection of the null hypothesis, indicating stationarity.

3.3.10 Cointegrated time series

Although stationarity is a key assumption of the econometric frameworks used in this research, it can be achieved not only through differencing (see Section 3.3.1). If the non--stationary time series are *cointegrated*, we can estimate one through the other despite their changing unstable properties.

Cointegration is a statistical property of two (or more) time series variables, that is used when the series are individually non-stationary but their linear combination is stationary. This means, that if Y_t and X_t are both non-stationary time series, they are cointegrated if there exists α such that $Y_t - \alpha X_t$ is a stationary series. Cointegration also means that the non-stationary parts of Y_t and X_t are related in such a way that their stochastic trends cancel each other, which ensures the stationarity of the residuals.

To detect cointegration, we use the Johansen test, which uses the vector error correction model (VECM), an extension of the vector autoregression model (VAR) modified to include cointegrated series. The Johansen test uses two types of statistics to examine the order of the coefficient matrix in the VECM that indicates the number of cointegration relationships: the trace statistic and the maximum eigenvalue statistic. Both are used to test the null hypothesis of the presence of cointegrating vectors.

3.3.11 Dealing with structural breaks

The period from the beginning of the measurement to around the middle of 2022 is characterised by a highly unusual growth rate (see Figure 3). Given the aftermath of the covid-19 pandemic and the outbreak of war in Ukraine, this period is very different from the rest of the time series. In these cases, it makes sense to split the dataset by a *structural break* and estimate the individual parts separately rather than trying to find a general equation for two disparate time periods.

To determine how the data should be partitioned, we use the *Chow test*, which can determine both the number of potential structural breaks and the timing of their occurrence.

The Chow test essentially compares the goodness of fit of one regression model applied to the entire data set with the goodness of fit of two models applied to two individual subsets of the data.

The test statistic for the Chow test is calculated as

$$F = \frac{(SSR_{full} - (SSR_1 + SSR_2))/p}{(SSR_1 + SSR_2)/(N - 2p - 1)}, \tag{3.21}$$

where SSR_{full} is the sum of squared residuals from the model fitted to the whole time series, SSR_1 and SSR_2 are the sums of squared residuals from the models fitted to each of the two subsets respectively, N represents a number of observations in the full dataset, p stands for the number of parameters estimated in each model.

The null hypothesis states, that the coefficients of the subsets are equal to the coefficients of the whole set, i.e., there is no structural break.

Chapter 4

Analysis of results

Weekly observations of the CMPI from October 2021 to June 2024 provide the basis for our analysis, encompassing a total of 144 distinct data points. First we present the results obtained from the CMPI measurements and compare them with the corresponding official indices. In the following part, we discuss the modelling of the CMPI by integrating the external variables and estimating the official price index (MIPI) using the CMPI as a predictive variable.

4.1 CMPI level over time

The findings of the CMPI analysis indicate that there was a 13.9% increase in the overall price level of construction materials over the 11 quarters assessed from October 2021 to June 2024. Despite the research spanning less than three years, the observed progression in the price level of this market segment encompasses a notably turbulent period (see Figure 3).

Following a period of consistent fluctuation around zero or near the 2% target set by the Czech National Bank (CNB) that lasted for several years, headline inflation in the Czech Republic began to accelerate in 2020, coinciding with the onset of the covid-19 pandemic's impacts. The same was true for producer prices, and therefore also for prices of building materials. The reverberations of the covid pandemic are evident in the sharp increase of CMPI observed from very beginning of measurement in 2021.

Nevertheless, the Russian invasion of Ukraine, initiated in February 2022, caused, among other consequences, shortages of certain building materials, and escalating energy prices. These factors have further intensified the already burgeoning growth in CMPI.

These effects were absorbed by the market during the second half of 2022, as the CMPI peaked on 7 August 2022 at 119.2%.*) Next, for three consecutive quarters, the CMPI values stabilized at a similar level around the basis index value of 118%. Subsequently, prices declined to below 115% starting from the second quarter of 2023. In the second quarter of 2024 – the final quarter of measurement – we observed a return to growth, though it was modest, not exceeding 2% on a quarter-on-quarter basis and even remaining negative on a year-on-year comparison.

Figure 3: Comparison of CMPI with official indices

October 2021 – June 2024, basis index (100% = 30 September 2021), official indices reported at month-end, percent

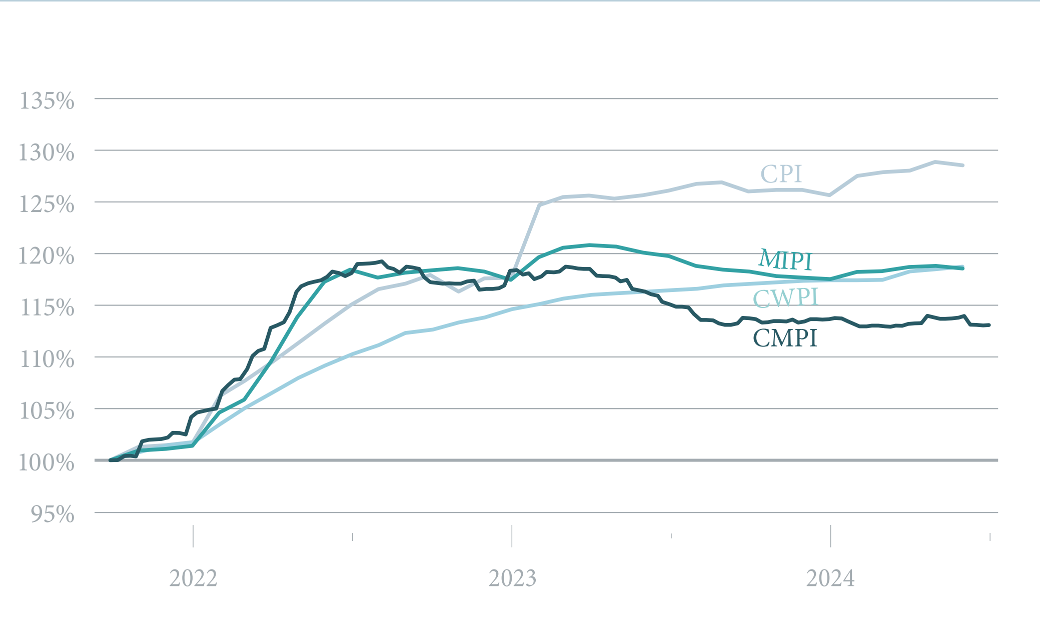


Figure 3 provides a summary comparison of CMPI with three other indices published by the CZSO: CPI – *Consumer Price Index*, CWPI – *Construction Work Price Index*, and MIPI – *Price Index of Material Inputs for Construction Work*. The MIPI (a complementary index to the CWPI) is most closely aligned with the CMPI, where there is a noticeable shift in these two indices from the beginning of 2023 due to the change in index schemes by the CZSO from January 2023.

Beginning in early 2023, the MIPI, for which CMPI should be a proxy, started to diverge from the CMPI following a period of over a year during which the two indices overlapped almost exactly. Nevertheless, the subsequent trajectories of the two indices remain quite similar, albeit with a noticeable gap. We attribute this divergence to changes implemented by the CZSO**) in its index schemes starting January 2023, which could lead to a shift in the price levels as measured by the official index.

*) The CMPI values in this thesis deviate slightly from those reported in previous work (Štefl, 2022) due to a revised methodology (see Section 3.1).

**) https://csu.gov.cz/docs/107508/4b88f8bc-e5a0-c4ca-a3a2-58269980a1d2/01104123q1m1cz.pdf

It should be noted that Figure 3 presents the individual values of the official indices as recorded at the end of each month, whereas the CMPI assigns these values to the actual date of measurement. As outlined in Section 3.1.2, these values ought to be correctly assigned to the middle of the month, as demonstrated in Figure 4. Following this adjustment, the month-on-month indices exhibit an average correlation of 0.92 with each other.

Figure 4: Month-on-month inflation measured by CMPI and MIPI

October 2021 – June 2024, month-on-month index, MIPI reported at mid-month, percent

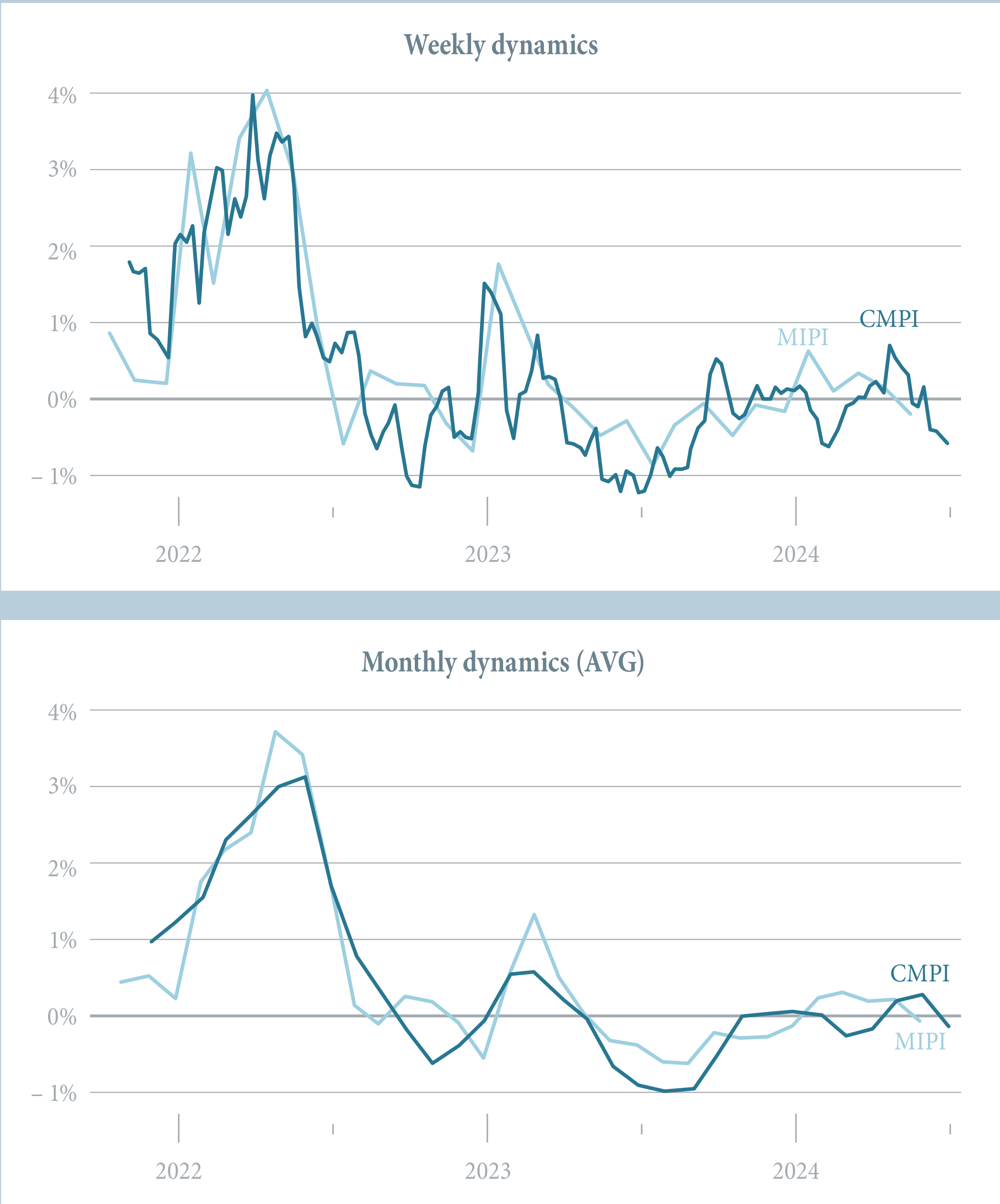


Figure 4 compares the month-on-month inflation of building materials as measured by CMPI with the official MIPI. The upper plot illustrates the weekly dynamics, shifting MIPI values to align with the midpoint of each month. The lower plot displays the monthly dynamics, focusing on the changes in the average monthly value of the basis index, smoothing their progression.

As of the date of publication of this work, the last measurement of the CMPI was made on 30 June 2024, and the last available figure for the official indices (MIPI and CWPI) was for May 2024.

4.1.1 Structural breaks due to pandemics and war

As discussed in Section 3.3.11, the first three quarters of the research are associated with an extraordinary increase in inflation, which is not only evident in the area of construction materials (see Figure 3). This time period is actually influenced by the aftermath of covid-19 pandemics and from February 2022 also by the Russian invasion of Ukraine and it is understandable to think of this era as of the period of the structural disruption.

Figure 5: Structural breaks within the Chow test

October 2021 – June 2024

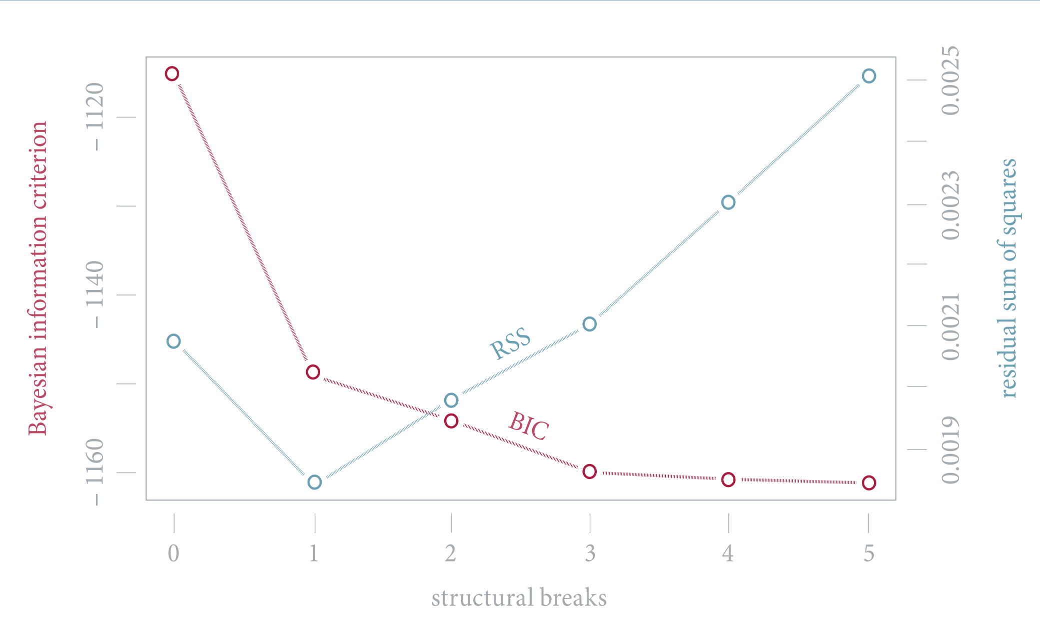


Figure 5 shows the result of the Chow test, which analyzes the CMPI time series for the presence of structural breaks. While the Bayesian information criterion (BIC) decreases as the number of structural breaks increases, indicating a better fit, the largest jump is associated with the inclusion of a single break. This is confirmed by the lowest value of the residual sum of squares (RSS) at the same point. The one structural break is also consistent with economic reasoning, as it is timed to 8 May 2022, when the effects of the covid-19 pandemic and the *immediate* shock caused by the Russian invasion of Ukraine passed.

This observation and conjecture is further supported by the result of the Chow test (see Figure 5), which indicates a reduction in BIC when breakpoints are included. Lower BIC values generally indicate better model fit, both with respect to model complexity and goodness of fit. The figure shows that adding one breakpoint reduces the BIC significantly more than adding additional structural breaks, where the effect is noticeably more modest.

Moreover, the lowest value of the residual sum of squares is reported when a single structural break is included and increases again with a higher number of breakpoints. In the case of the week-on-week data, this breakout falls on 8 May 2022 (see Figure 6), suggesting that the construction market has been able to deal with the consequences of the conflict in Ukraine within less than a quarter.

Figure 6: Week-on-week CMPI with one structural break

October 2021 – June 2024, week-on-week index, percent

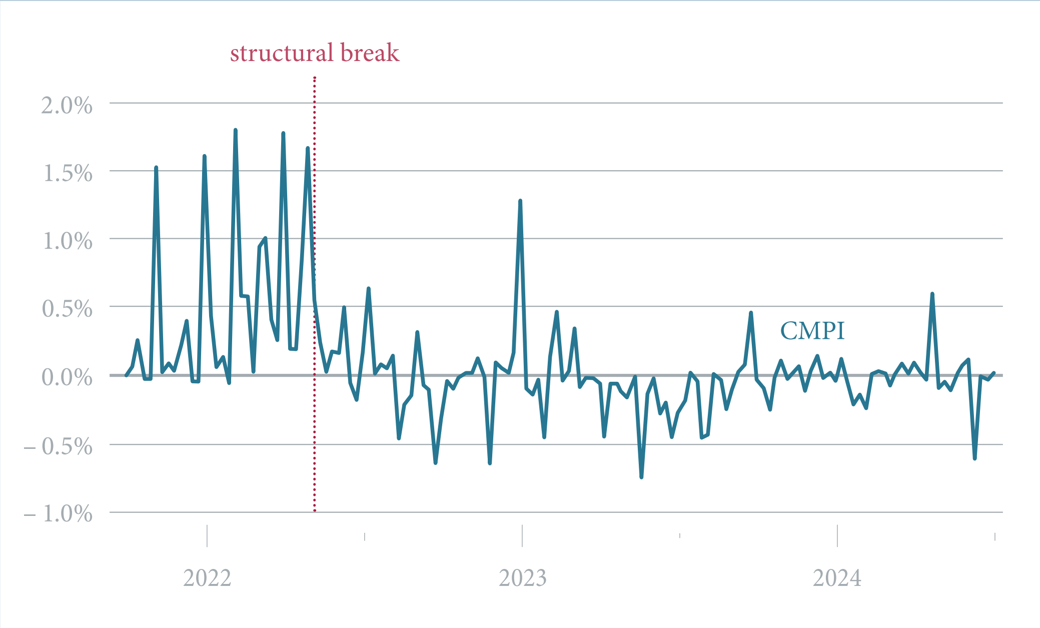


Figure 6 shows the week-on-week CMPI data with a structural break based on the Chow test. The structural break falls on 8 May 2022, less than a quarter after the Russian invasion of Ukraine. While there are clearly higher weekly increments in the time period prior to the break, which are related to the inflationary period following the covid-19 pandemic and the effects of the conflict in Ukraine, the remaining time series – while showing few disproportionate movements – is stationary and generally exhibits different statistical characteristics.

Given the small number of observations (31 weeks) in the period before the structural break, which in our opinion does not allow us to build a reliable ARIMA model, we consider two intervals for model building: from the breakpoint to the end of June 2024 and the entire span of measurements (i.e., without taking the structural break into account).

4.1.2 Seasonality and trend

We did not find any major evidence of seasonality in the weekly data. From a theoretical point of view, there are several possible causes for seasonality in inflation data, e.g. a change in price lists associated with the beginning of the year, price adjustments related to the beginning or end of the construction season, etc. In addition, the headline CPI inflation in the Czech Republic shows a strong seasonal pattern at the month-to-month level (Franta & Sútoris, 2022). It would therefore be appropriate to include a seasonal component in the ARIMA model.

However, given the less than three-year span that our research covers, we believe that it would be inappropriate to draw strong conclusions about seasonality. Moreover, including the seasonality term did not improve the model fit in any of tested combinations. Thus, overall, we primarily only remove the trend from the data through first-order differencing. Hence, we achieve stationarity in the model inputs, meeting a fundamental prerequisite for the use of the chosen econometric frameworks.

4.1.3 CMPI decomposition

As previously noted, the high level of granularity in the underlying CMPI data enables the construction of specific sub-indices with up to four distinct tiers. This detailed underlying structure of the index easily allows for more accurate use of the data and deepens our understanding of market price movements. A summary of the evolution of the selected major components of the CMPI compared to the behaviour of the overall index is presented in Figure 7.

For illustrative purposes, we have selected four subcategories that play the most significant role in the overall CMPI, as altogether they account for almost 50% of the overall CMPI basket structure. *Cement & Aggregates* represent approximately 15% of the total cost of the CMPI consumption basket, both *Masonry* and *Insulation & Fabrics* account for roughly 13%, and *Metalworks* stands for around 7% of the total costs. The results show that *Masonry, Insulation & Fabrics* and *Cement & Aggregates* mostly increased the overall CMPI value, while *Metalworks* drove its value down (see Figure 7).

The price of *Masonry* was clearly already rising at an above-average rate even before the outbreak of the Russian invasion of Ukraine (which was initiated on 24 February 2022), which we attribute to the still-present consequences of the covid-19 pandemic.*) After the war conflict began, its growth accelerated further, peaking at 132.5% on 4 September 2022, and returning to around the average level of the overall CMPI during the second half of 2023.

Insulation & Fabrics experienced the most significant increase only after the onset of the war in Ukraine, prior to this, their growth rates remained at average levels. Prices in this group peaked (134.6%) slightly earlier than *Masonry* on 17 July 2023, since when they have gradually converged back to the CMPI average. The second quarter of 2024 showed

*) The last restrictive measures related to covid-19 in the Czech Republic ended in spring 2022.

a more significant price increase, which may be linked to the start of the construction season. However, this shift is not noticeable in previous years, so we do not dare to label this fluctuation as a sign of seasonality without a longer-term examination.

The price level in the *Cement & Aggregates* group grows most steadily among the selected sub-indices and shows no greater sensitivity to external shocks than is common for the overall average. Its value reached a maximum (127.5%) on 31 December 2023. We attribute the characteristics of the development in this group to the fact that the production of a substantial part of the items takes place domestically, including the exploitation of primary raw materials.

Figure 7: Results of CMPI sub-indices

October 2021 – June 2024, percent

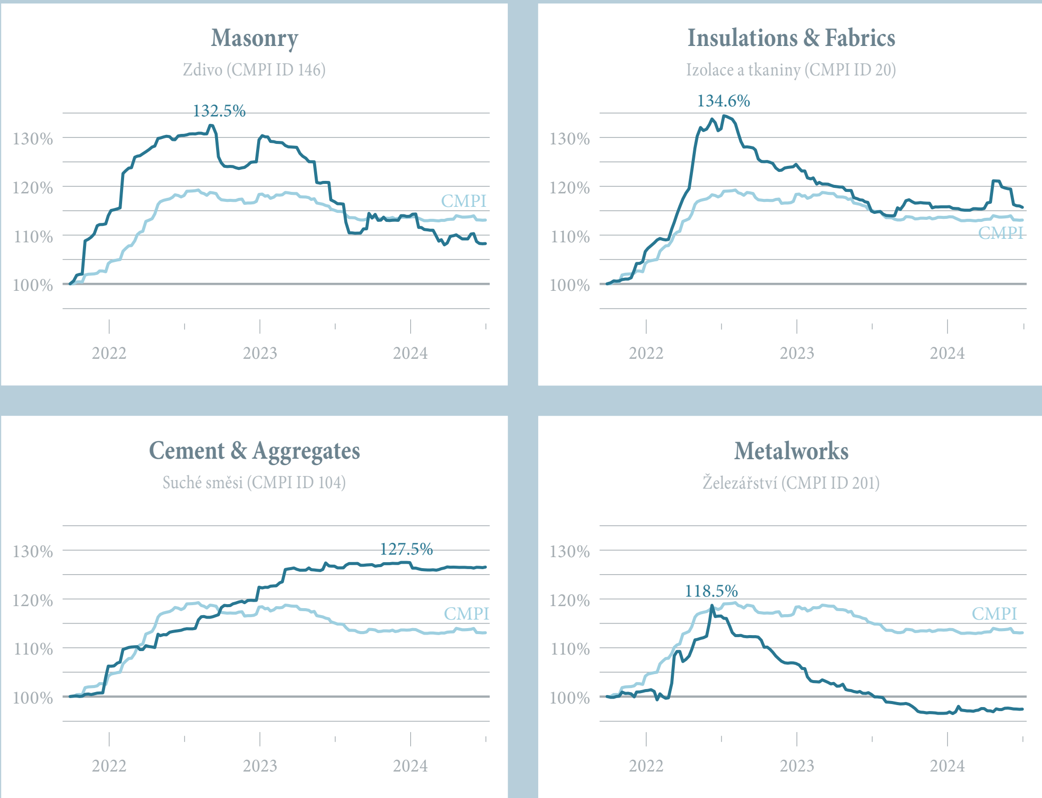


Figure 7 summarises in detail the movements in the price level of the selected significant building materials through the relevant sub-indices and compares them with the evolution of the overall CMPI. With the exception of *Metalworks*, all categories correspond to the *Product Group Level 2* of the CMPI directory table.

On the other hand, items within the *Metalworks* group exhibited a very strong reaction to the outbreak of war in Ukraine. This strong response is unsurprising, considering that a predominant origin of production raw materials in this group from Ukraine and Russia. Apart from its peak (118.5%) on 12 June 2022, the value of the corresponding sub-index has consistently remained below the overall CMPI, and as of August 2023, the average price level in this category is even lower than it was before the measurement started in October 2021.

The effects of the war in Ukraine, which are most evident in *Insulation & Fabrics* and *Metalworks* among the selected groups, were also manifested in the period immediately after 24 February 2022 by the temporary unavailability of a considerable number of the corresponding items across all eshops under review. The most pronounced effect of the shortage of building materials was seen in facade polystyrene, which was not in stock in large quantities until the middle of 2022 (Štefl, 2022).

4.2 Estimating CMPI through commodity prices

Before estimating the official MIPI through the CMPI, we develop a model to estimate the CMPI using commodity prices. As discussed in detail in Section 3.3, the primary framework we use for this purpose is the ARIMA model.

First, we focus on building the ARIMAX model from the selection of appropriate variables to the correct setting of ARIMA parameters, then we build the DFM approach. Together with other benchmark models, we then subject all approaches to a comparative scrutiny.

Given that commodity data are available at a daily frequency, all models specified in this section estimate the CMPI at the highest possible frequency, i.e., weekly. The commodity data are further transformed to match the same time distribution.

4.2.1 Granger causality

We begin the selection of exogenous variables for the ARIMAX model with the Granger causality test (see Section 3.3.4). It is important to note that although we examine only eleven variables, the number of test inputs increases significantly with their different variations (such as their moving averages or other intertemporal representations) and lags.

Indeed, in many cases the price from the commodity exchanges may not be an immediate determinant of the cost of production inputs, just as the overnight extraordinary fluctuations that naturally occur in high-frequency prices from exchanges (see Table 4) may seep into producers' long-term contracts in a much more muted form.

As the results of the Granger causality test indicate (see Table 5), this is particularly true for energy commodities - natural gas and electricity. Of all the proxies tested, their month-on-month representation is found to have the largest effect on the predictability of the week-on-week CMPI in the energy commodity group.

The test shows that the most influential variable in our sample is the one-week lagged iron ore price. In terms of true causality, this seems odd, as we doubt that this commodity actually enters the production process of any building material in this crude form. We would have expected it not to outweigh the impact of steel, which is much closer to the final products, but it turns out that the three-week lagged price of steel is second only.

Table 5: Results of Granger causality test

October 2021 – June 2024, weekly frequency (CMPI WoW), MA – Moving Average, W – Week, M – Month

Commodity	Smoothing	Basis	Lag	*p*-value
Iron Ore		WoW	1 week	0.0002
Steel		WoW	3 weeks	0.0006
Natural Gas		MoM		0.0021
Aluminum		WoW	3 weeks	0.0024
Diesel	MA (week)	WoW		0.0027
Petrol	MA (week)	WoW	3 weeks	0.0067
Electricity		MoM		0.0077
Crude Oil_{WTI}		WoW	1 month	0.0079
Lumber	MA (week)	WoW	2 months	0.0228
Crude Oil_{Brent}		WoW	1 month	0.0298
Copper	MA (week)	WoW		0.1895

Table 5 summarizes the p-values of the Granger causality test for the 11 variables in relation to the CMPI. For each commodity, only the best-fit representation is reported. At the 5% significance level, Granger causality suggests that *Copper* price does not help in predicting the CMPI.

Some variables are strongly pairwise correlated, namely *Crude* Oil_{WTI} and *Crude* Oil_{Brent} (correlation 0.97), *Natural Gas* and *Electricity* (0.94). To avoid overfitting, we select only those with lower *p*-values from the Granger causality test from these pairs for further analysis (i.e., *Crude* Oil_{WTI} and *Natural Gas*). Similarly, *Petrol* and *Diesel* are also correlated (0.88), but following Granger causality, we select the three-week lagged diesel price, which is correlated with the unlagged gasoline price with a very low correlation coefficient of 0.04. Thus, this pair can be viewed as if we only included one of them in the model, but twice – lagged and unlagged – which is a valid approach. Moreover, Granger causality is not the only procedure by which we select exogenous variables.

In addition to the variables tested in Table 5, we also examined the incluence of CNB interest rate changes. We dealt with it separately because it is inherently very different from the commodity market data. The Granger causality analysis showed that a two-month lagged change in the interest rate helps predict the weekly CMPI more than the majority of commodity prices (p-value 0.0006). However, its addition to the ARIMAX model significantly reduced the significance of individual variables (it was not significant even on its own) and the quality of the model as a whole, so we decided to disregard it.

We subjected individual variables to the same examination for the period after the structural break, but the results differed considerably from those presented in Table 5 (see Appendix B). Given the smaller number of observations in this post-break period, we took a less dogmatic approach to the results for subsequent ARIMAX model building, preferring also to consider the results relative to the whole period.

4.2.2 LASSO regression

Another way to reduce the number of variables is LASSO regression (see Section 3.3.6). We apply it on two datasets – on all variables including their different representations, smoothing and lags (see Table 6), and on a subset selected according to Granger causality reported in Table 5 (including *Lumber* and *Copper* to cover all commodities).

Table 6: Results of LASSO regression

October 2021 – June 2024, weekly frequency (CMPI WoW), MA – Moving Average, W – Week

Commodity	Smoothing	Basis	Lag	LASSO coefficient	Granger p-value
Iron Ore	MA (week)	WoW	3 weeks	0.0085	0.0087
Petrol	MA (week)	WoW	3 weeks	0.0028	0.0351
Natural Gas	MA (week)	WoW	1 week	0.0022	0.0201
Electricity	MA (week)	WoW	1 week	0.0018	0.0239
Diesel	MA (week)	WoW	3 weeks	0.0013	0.0084
Crude Oil$_{Brent}$	MA (week)	WoW	1 week	0.0009	0.8425
Copper	MA (week)	WoW	3 weeks	0.0005	0.5725

Table 6 reports the coefficients from the LASSO regression ($\lambda = 0.0006$) and the associated p-values from the Granger causality test. *Iron Ore* shows the highest coefficient, which, along with *Petrol*, *Natural Gas*, *Electricity* and *Diesel*, helps predict the CMPI according to Granger causality. Interestingly, this method only selects average values on a week-on-week basis.

For the full list of variables, using the cross-validation method, we find that the optimal regularization parameter λ^* is equal to 0.001 (see Section 3.3.6). However, at this level of λ, the penalty term is too high to select any variable. Therefore, we reduce the penalization by multiplying λ^* by 0.8, 0.6, 0.4 and 0.2, and then check the quality of the resulting ARIMAX models.

For $\lambda = 0.0008$, the ARIMAX model consisting of variables distilled from LASSO regressions shows both the lowest AIC and the highest number of significant variables (*Copper*, *Crude Oil*$_{WTI}$, *Diesel*). The results of the LASSO regression, ordered by the magnitude of the coefficient, are presented in Table 6.

Applying the LASSO regression ($\lambda = 0.6\,\lambda^*$) to the subset consisting of only the variables selected by the Granger causality test reconfirms the importance of *Iron Ore*, *Petrol* and *Electricity* (all with the same specifications as in Table 6).

4.2.3 ARIMAX model

It is important to note that neither the Granger causality results nor the LASSO regressions give us a strict criterion for deciding which external variables to include in the ARIMAX model. The relationships between variables in time series can be complex and predictors should be selected taking into account all relevant information, including the domain knowledge. The same is true for setting up ARIMA components.

Influential exogenous variables

We performed Granger causality test, LASSO regression and tested the fitting of the model with external variables in multiple representations:

- smoothing: no, MA (week), MA (14 days), MA (30 days);
- basis: week-on-week (WoW), month-on-month (MoM);
- lag: no, 1 week, 2 weeks, 3 weeks, 1 month, 2 months, 3 months.

All models where the p-value of the exogenous variable is greater than 0.1 are disregarded. According to these criteria, following variables were found the most influential:

- *Diesel* (MA week);
- *Electricity* (MA week, 1 week lag);
- *Crude Oil*$_{WTI}$ (MA week, 2 months lag);
- *Copper* (MA week, 1 week lag).

Although the LASSO regression favours *Petrol* over *Diesel*, based on Granger causality and the fact that diesel fuel is used for industrial transport rather than petrol, we opt for *Diesel*. Moreover, both variables are highly correlated. Omitting *Petrol*, the LASSO regression selected *Diesel* (MA week, 3 weeks lag).

Due to the high energy intensity of building materials production, we consider it appropriate to include an energy commodity – *Electricity* or *Natural Gas* – in the model. Due to their correlation, we decided to select only one of them. Despite the results of Granger causality and the LASSO regression, we decided to lean towards *Electricity* (MA week, 1 week lag). We believe that in addition to the clear impacts on production costs, electricity has a more significant impact on retailers' costs compared to natural gas.

Although the *Crude Oil*$_{WTI}$ may have an effect on the price of plastic products, thereby feeding into the cost of producing building materials, it was not significant in most of the models built and so we omit it from the final selection.

After initial analysis, *Copper* (MA week) appears to not Granger-cause CMPI changes. In terms of construction costs, precious metals do not play a major role, especially compared to other commodities. However, the LASSO regression selects copper both if we use all variables, but also if it is run only on the sample of commodities previously selected by the Granger causality test. Furthermore, in most models, it achieves significance greater than 99% and also improves their forecasting performance, so we decided to keep it in the sample, although we are hesitant about its causality in relation to CMPI.

The main argument for this choice is that the model including these three variables shows the lowest BIC value of all tested models including models accounting for seasonality.

ARIMA model settings

Autocorrelation functions (ACF) and partial autocorrelation functions (PACF) are essential tools in time series analysis, especially for identifying appropriate ARIMA components p, d, and q (see Section 3.3.1). Based on a thorough analysis of the ACF and PACF results on both the full set of data points and the subset after the structural break (see Appendix A), we constrain the autoregressive component $p \in \{0, 1\}$, the differencing degree $d \in \{0, 1\}$ and the moving average component $q \in \{1, 2\}$. We estimate models for all combinations of these parameters and compare which one best fits the data according to the metrics defined in Section 3.3.7.

The best combination of AIC, BIC and RMSE values are achieved for the setting $p = 1$, $d = 0$, and $q = 1$. It is worth noting that this setting works well also for models with different selection of exogenous variables.

Benchmark models

In order to evaluate the models, we create several benchmark models against which we then compare ours. Although the AR(1) model is very simple, it would be all the more conclusive if our models did not outperform it. Moreover, the data show signs of autocorrelation, so this naive view is justified.

Next, we use an ARIMA model (excluding the exogenous variables) with the same settings as our best ARIMAX model (with $p = 1$, $d = 0$, and $q = 1$), which should help us clearly decide whether external regressors help in predicting CMPI.

Dynamic factor model

We also construct the ARIMAX model with the dynamic factors as external regressors. Using the shortened list of variables from Granger regression and the DFM, we obtain a set of three common factors (see Section 3.3.3), by which we effectively decrease the dimensionality of our model.

The ARIMAX model with $p = 2$, $d = 0$, $q = 1$, including two factors, showed the lowest AIC and BIC values. Nevertheless, the individual factors were not significant at more than 90% in any combination of the underlying variables, suggesting that the DFM approach does not perform well compared to direct variable selection.

ARIMAX evaluation

Comparison with benchmark models shows that the performance of all approaches appears to be relatively similar. However, the lowest RMSE and AIC is associated with the ARIMAX model including three exogenous variables. Despite the slightly lower BIC for the standard ARIMA, the inclusion of commodity prices seems to help estimate CMPI.

Table 7: Fit evaluation of ARIMAX model with commodity prices

October 2021 – June 2024, weekly frequency (CMPI WoW, differenced), percentage points (RMSE, MAE)

Model	AIC	AICc	BIC	RMSE	MAE
ARIMAX	−1178.90	−1178.28	−1158.21	0.356	0.224
$\text{ARIMAX}_{\text{DFM}}$	−1168.01	−1167.57	−1150.28	0.378	0.243
ARIMA	−1171.38	−1171.21	−1159.56	0.379	0.247
AR(1)	−1121.01	−1120.92	−1112.14	0.457	0.315

Table 7 shows the comparison of two developed models with their benchmarks for Akaike information criterium (AIC), Bayes information criterion (BIC), root mean square error (RMSE) and mean absolute error (MAE). By most metrics, the ARIMAX model with commodity prices outperforms its benchmarks, but the small difference is not very conclusive.

On the other hand, from a comparison of the fitting of the model and its benchmark, the addition of exogenous variables does not appear to be as strong, suggesting that building materials inflation is more dependent on other factors such as consumer sentiment.

This implication is supported by the model on restricted data (after the structural break) not performing significantly better than benchmarks, and the external variables being insignificant. From comparing the models, we conclude that the model with external variables can indicate exceptional periods, such as after the Russian invasion of Ukraine. However, during normal times, it performs comparably to the classical ARIMA model.

4.2.4 Forecasts of CMPI

The conclusions from the previous section that the addition of exogenous variables does not provide an improvement over the ARIMA model with the same settings are clearly confirmed by cross-validation (see Section 3.3.8). It shows that the model with commodity prices does not predict CMPI better in the long run than the model without them.

Figure 8: Cross-validation error metrics for one-week CMPI forecasts

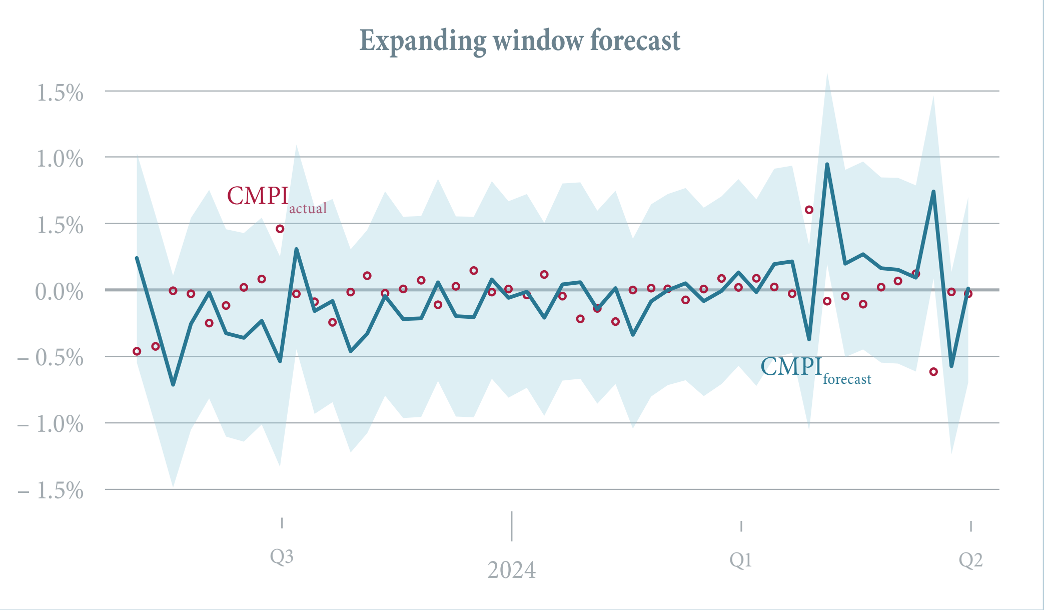
Weekly frequency (CMPI WoW), percentage points

Figure 8 compares the errors in one-week-ahead CMPI predictions using the model with and without external variables for different lengths of the training dataset. The ARIMA without commodity prices achieves higher predictive accuracy in both expanding and rolling window forecasts.

To give an idea of how the one-week CMPI forecast via ARIMAX with commodity prices matches the actual measured data, we report the comparison in Figure 9. In the case of the expanding forecast window, the training dataset starts fixed with the start of the research, in October 2021, and its end shifts from July 2023 to June 2024. The training dataset of the rolling window forecast always covers 95 weeks (i.e., about two-thirds of all

Figure 9: Comparison of one-week forecasts and actual data for CMPI

July 2023 – June 2024, weekly frequency (CMPI WoW), percent



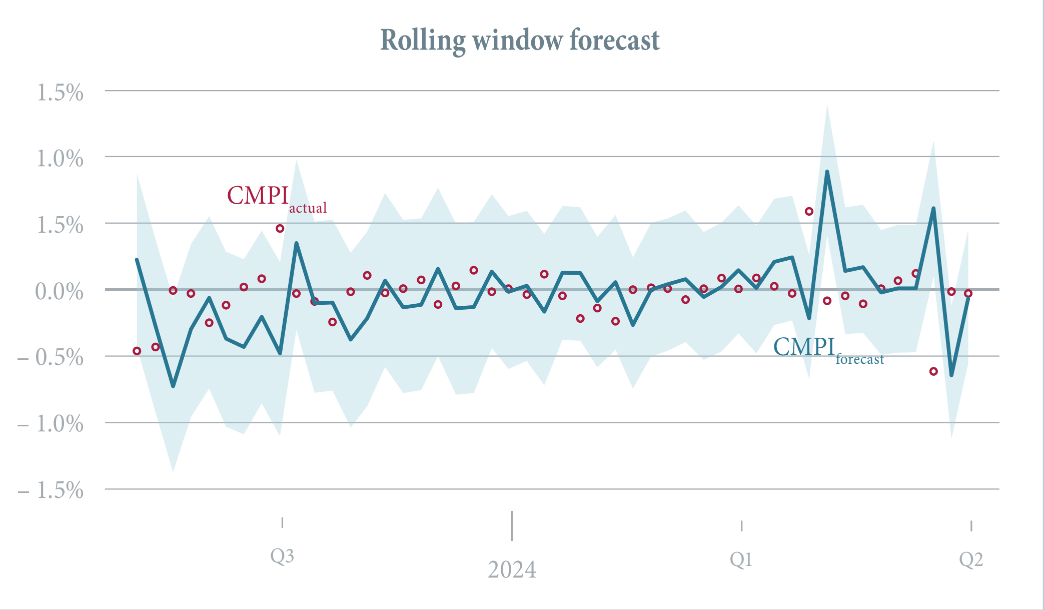


Figure 9 shows one-week-ahead CMPI forecasts using the ARIMAX model with commodity prices, their 95% confidence intervals, and actual CMPI values. We validate the week-on-week data on one-third of the observations (i.e., 48 weeks), so the size of the rolling window is 95 weeks.

observations), i.e., from February 2024 onwards it does not include the period before the structural break. Its accuracy, however, remains essentially unchanged. Despite the fact that most of the actual points are within the 95% confidence interval (see Figure 9), it is clear that in short-term estimates the inclusion of commodities in the ARIMA model does not help its predictive power (see Figure 8).

4.2.5 Estimating the *Metalworks* sub-index

Since we have not been able to demonstrate a clear relationship between the value of the overall CMPI and commodity prices, following Section 4.1.3, we use the granularity of the underlying CMPI data to test whether this link holds at least at the sub-index level. To test this hypothesis, we selected the *Metalworks* sub-index, which exclusively includes items composed of metal materials. Intuitively, one would expect the retail price within this sub--index to be primarily influenced by fluctuations in the prices of raw metal commodities.

We follow the same strategy as for fitting the overall CMPI, i.e., from selecting the relevant variables and their correct representations, to setting up the ARIMAX model. Finally, we verify the accuracy of the estimates of this sub-index in a pseudo-real time experiment.

Granger causality

Granger causality analysis indicates that the price of *Steel* has the most significant influence on the selected sub-index, with only *Diesel* prices also surpassing the 5% significance level. In contrast, the prices of *Iron Ore*, *Copper*, *Aluminum*, and energy commodities (*Electricity* and *Natural Gas*) do not exceed this threshold. The analysis also shows that for the variable *Steel*, the most important determinant is its average value in week-on-week representation, while for the price of *Diesel* it is the monthly average.

LASSO regression

If we run a LASSO regression on all metal commodities in our data (*Aluminum*, *Copper*, *Iron Ore*, *Steel*), the price of *Natural Gas* and *Diesel* (as in the previous section, we do not include *Electricity* and *Petrol* because of their correlation with *Natural Gas* and *Diesel*), we confirm the importance of *Steel*, *Diesel* and also the *Iron Ore* (for *Iron Ore* and *Steel* the correlation is only 0.21). Thus, we select these variables as exogenous regressors for the subsequent ARIMAX model.

ARIMA model settings

As in the case of estimating the CMPI, from the ACF and PACF plots and subsequent analysis of goodness-of-fit measures (see Section 4.2.3), we determine the best-fit settings of the ARIMA model as $p = 1$, $q = 0$, $d = 1$. In this setting, all selected variables are significant at more than 90%.

Benchmark models

We select benchmark models following the same approach as in the previous case (see Section 4.2.3), so we include a simple AR(1) model and an ARIMA model without exogenous variables with the same settings as in the ARIMAX model.

Dynamic factor model

In order to capture the underlying patterns in the related variable, we develop the dynamic factor model also for the *Metalworks* sub-index. As the input variables, we use the same list as for the LASSO regression (*Aluminum*, *Copper*, *Iron Ore*, *Steel, Natural Gas, Diesel*). The best fit is achieved for one factor with the same ARIMA settings as for the the ARIMAX model ($p = 1$, $q = 0$, $d = 1$). In this case, all components, including the factor, are significant at 90% or more.

ARIMAX evaluation

According to the information criteria and error measures, the fit of the ARIMAX model for the Metalworks index with exogenous variables is better than that of all its benchmark models. This difference appears to be more convincing than for the overall CMPI index.

Table 8: Fit evaluation of ARIMAX model for *Metalworks* sub-index

October 2021 – June 2024, weekly frequency (CMPI WoW, differenced), percentage points (RMSE, MAE)

Model	AIC	AICc	BIC	RMSE	MAE
ARIMAX	−967.11	−966.49	−946.42	0.752	0.431
$\text{ARIMAX}_{\text{DFM}}$	−964.68	−964.39	−949.90	0.769	0.419
ARIMA	−955.41	−955.23	−943.58	0.801	0.432
AR(1)	−912.36	−912.27	−903.49	0.953	0.534

Table 8 shows the comparison of two developed models for the *Metalworks* sub-index with their benchmarks for Akaike information criterium (AIC), Bayes information criterion (BIC), root mean square error (RMSE) and mean absolute error (MAE). The ARIMAX model with commodity prices outperforms its benchmarks in all metrics.

However, a comparison of the error metrics with those obtained from fitting the overall CMPI (see Table 8) reveals an increase in their values for the *Metalworks* sub-index. This result is contrary to expectations, as the narrower focus should yield lower error metrics.

Forecasting performance

The cross-validation of the models for nowcasting the metalworks sub-index shows that the model with exogenous variables fails to consistently outperform the corresponding benchmark without exogenous variables even in this case. While the difference in estimation accuracy is noticeably smaller (see Figure 10) than in the case of nowcasting the overall CMPI (compare with Figure 8), it can be concluded that the model for estimating the *Metalworks* sub-index using commodity prices is at best as good as the corresponding ARIMA, but does not clearly outperform it. Hence, despite the intuitive expectation, prices of metal commodities do not enhance the forecast of the related sub-index.

Figure 10: Cross-validation of one-week *Metalworks* sub-index forecasts

Weekly frequency (CMPI *Metalworks* sub-index WoW), percentage points

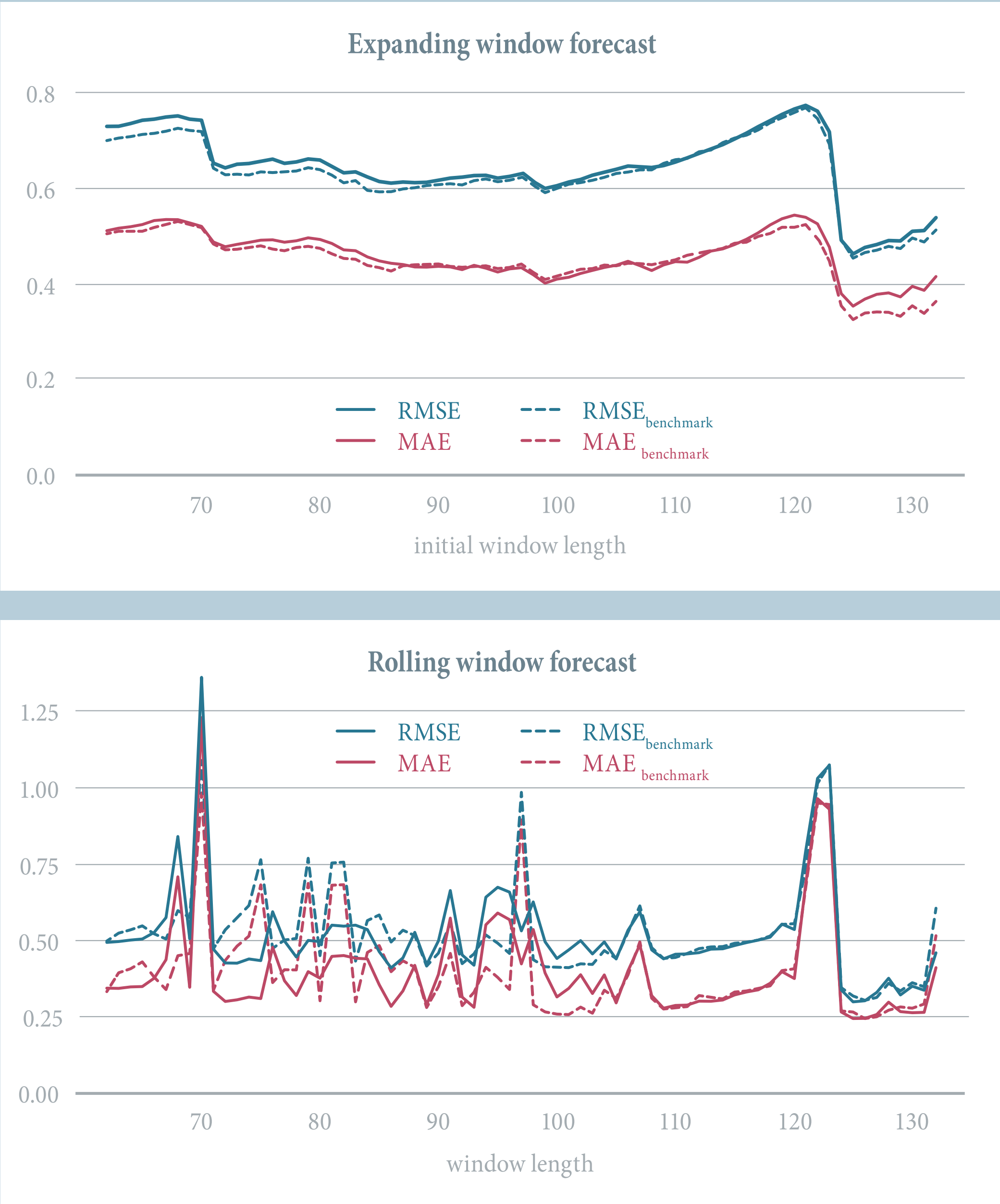


Figure 10 compares the errors in one-week-ahead predictions of the *Metalworks* sub-index using the model with and without external variables for different lengths of the training dataset. The ARIMA without commodity prices achieves higher predictive accuracy in both expanding and rolling window forecasts.

This finding further confirms the conclusion from the previous section, namely that commodity prices on world exchanges do not have a direct impact on retail prices of building materials in the Czech Republic. Given that even in the specific case where the greatest interconnectedness was anticipated – estimating a subindex for *Metalworks* through metal commodity prices – this was not true, we find this conclusion compelling.

4.3 Estimating official MIPI through CMPI

The greatest advantage of our CMPI, which is based on web-scraped data, lies in its almost immediate availability following data collection. This rapid turnaround allows the CMPI to be published approximately one month before the official MIPI (see Section 3.1.2). The prompt availability of the CMPI thus directly encourages the nowcasting of the official MIPI through the CMPI.

4.3.1 ARIMAX model

As in Section 4.2, we also base the nowcasting of MIPI on the ARIMAX framework, where CMPI enters as an exogenous variable. After testing various transformations, the average month-on-month evolution of the CMPI appears to be the most useful, so we choose it as the only external variable.

Cointegration of CMPI and MIPI

As was the case with the weekly data, without additional transformations, both variables are non-stationary on a monthly basis. However, according to the results of the Johansen cointegration test, these series share a long-run equilibrium relationship. In particular, the Johansen test shows that there is at least one cointegration vector between the two series, which means that their linear combination leads to a stationary series (see Section 3.3.10). The finding of cointegration is crucial because we can include these variables directly in our models without differencing or further adjustments.

ARIMAX model settings

In the same way as in the case of estimating the CMPI using commodity prices (see Section 4.2.3), we first determine the possible settings of each ARIMA component from the ACF and PACF analysis and then algorithmically select the model with the best values of AIC, BIC and RMSE.

The best values of these metrics are achieved for the combination $p = 2$, $d = 0$, and $q = 1$, which is the same as in the case of estimating the CMPI. At this setting, the regressors are also significant at more than 99%.

Benchmark models

For model evaluation purposes, we again use several benchmark models. The simplest one is again the AR(1) model, which serves more as a general validation of the approach. More interesting is the ARIMA built on MIPI only, where in this case, however, we must first differentiate the series to achieve stationarity. The last benchmark is a simple OLS model based on the linear regression.

We also use these models to cross-validate the predictions. Given the high correlation between the two series and the one-month time lag ahead of the official index, we also add to these the last benchmark, which is simply the CMPI value at the corresponding time.

ARIMAX evaluation

The comparison of the models clearly shows that the inclusion of CMPI as a regressor to model the official MIPI significantly improves both the quality of the model and its fitting accuracy (see Table 9). The model built on the ARIMAX framework performs best of all, although OLS does not perform poorly either.

Table 9: Fit evaluation of ARIMAX model with CMPI

October 2021 – June 2024, monthly frequency (MIPI ~ CMPI), percentage points (RMSE, MAE)

Model	AIC	AICc	BIC	RMSE	MAE
$\text{ARIMAX}_{\text{CMPI}}$	−227.74	−226.26	−220.41	0.568	0.460
OLS_{CMPI}	−225.15	−224.73	−220.75	0.653	0.513
ARIMA	−193.06	−192.65	−188.76	0.947	0.675
AR(1)	−192.47	−192.33	−189.60	1.001	0.676

Table 9 shows a comparison of the models for fitting MIPI based on CMPI ($\text{ARIMAX}_{\text{CMPI}}$, OLS_{CMPI}) with their benchmarks for Akaike information criterium (AIC), Bayes information criterion (BIC), root mean square error (RMSE) and mean absolute error (MAE). Due to the noticeably better performance of models incorporating CMPI, we conclude that this index significantly improves both the overall quality of the model and its accuracy in fitting MIPI.

If we compare the ARIMAX model including CMPI with the same ARIMA model, we find that the fit accuracy can be improved by nearly 40% (in terms of RMSE) if we include CMPI. In this case, the error measured by RMSE is 0.57 p.p. over the whole sample.

It is also worth noting that, when it comes to AIC, AICc and BIC, the AR(1) model performs nealy the same as the ARIMA model including the components of differencing and moving average. However, its accuracy is worse, albeit slightly. We attribute the similarity of the models mainly to the fact that we had to differentiate the data for the AR(1) model in advance, so the only extra component that ARIMA contains is the moving average.

4.3.2 Forecasts of MIPI

Our findings from model fitting presented in the previous section are confirmed in the pseudo-real time experiment, where models using the CMPI can consistently estimate the MIPI value one month ahead better than models that do not use it.*) Thus, it can be concluded that the CMPI-based models can nowcast the official MIPI with an accuracy of approximately 0.35 p.p., i.e., roughly 25% more accurate than the reference models.

Table 10: Cross-validation of MIPI one-month ahead forecasts

March 2023 – June 2024, monthly frequency (MIPI ~ CMPI), percentage points

	Expanding window		Rolling window	
Model	**RMSE**	**MAE**	**RMSE**	**MAE**
$ARIMAX_{CMPI}$	0.349	0.308	0.332	0.289
OLS_{CMPI}	0.361	0.307	0.363	0.289
CMPI (direct)	0.377	0.308	0.377	0.308
ARIMA	0.478	0.380	0.479	0.384
AR(1)	0.530	0.431	0.437	0.366

Table 10 provides a comparison of approaches to MIPI prediction based on the accuracy of its one-month ahead forecasts over the expanding and rolling window. The CMPI-based models ($ARIMAX_{CMPI}$, OLS_{CMPI}) perform considerably better than their benchmarks without CMPI (ARIMA, AR(1)). Direct estimation via unmodelled CMPI also outperforms these benchmarks, but does not perform better than models that employ it as the exogenous variable. For the expanding window, which starts with 16 months of data, ARIMAX provides the most accurate estimates. For the rolling window, which captures short-term changes over a 10-month period, the results are the same.

In the case of an expanding window that starts with 16 months of data (i.e., half of all measurements), the best model is ARIMAX with CMPI as an exogenous variable, where the RMSE is 0.35 p.p. (see Table 10).

As for the rolling window forecast, where the model is trained on the preceding 12 months of data, an OLS regression provides an even better result than ARIMAX, where the average RMSE over the 16-month test period is slightly more than 0.33 p.p.

For the purpose of nowcasting, i.e., instantaneous estimation, rolling window is generally a more appropriate method as it captures more short-term patterns. The training period of 10 months was not chosen randomly, but after testing all lengths from 6 to 24 months with 10 months, the most accurate estimates emerged.

*) The ARIMA and AR(1) benchmark models were estimated on differenced series to achieve stationarity. To make the forecasts comparable, we subsequently integrate the series estimates.

In our case, the advantage of the moving window forecast is that it stops accounting for the period before the structural break with a delay. Figure 11 graphically proves that in the last estimated quarter, when these disturbing data are long gone from the training period, the accuracy of the rolling window estimation slightly improves relative to the respective expanding window forecast.

Figure 11: Comparison of MIPI nowcasts and the actual data

March 2023 – June 2024, monthly frequency (MIPI MoM), ARIMAX (MIPI ~ CMPI), percent

Figure 11 shows nowcasts of MIPI using the ARIMAX model with CMPI, their 95% confidence intervals, and actual MIPI values. We validate the month-on-month data on a half of the observations (i.e., 16 months), the size of the rolling window is 10 months.

4.4 Research limitations

We conclude the results section by outlining potential limitations of our research, accompanied by a brief commentary and discussion. We also offer recommendations for future research possibilities.

- **Short duration**: The research spans slightly less than three years, a relatively short period to capture all possible economic cycles and longer-term trends. For this reason, we focus on short-term estimation in this paper, but this duration may also limit the ability to generalise the findings on the positive effects on the accuracy of nowcasting, as it was only validated for 16 months of testing.
- **Modeling constraints**: Although ARIMA models are generally suitable for modelling time series, they assume, among other things, that its parameters do not change over time. However, this assumptions may not always hold, and it may be preferable to use more flexible time-varying parameters models, such as the score driven models.
- **Generalizability of findings**: The research focuses on the Czech construction materials market, which may limit the generalizability of the findings to other markets or sectors. The specific economic conditions or market dynamics in the Czech Republic might not be directly comparable to other contexts.
- **External factors**: The period studied includes specific external events like the recovery from covid-19 pandemic and geopolitical tensions (the Russian invasion of Ukraine), which clearly caused atypical market behaviors, which we have tried to treat with the structural break approach. However, given the apparent trade-off between dataset size and consistency, it remains plausible that we have not adequately mitigated the effects of discovered structural break.
- **Economic complexity**: Given that inflation cannot usually be explained by simplistic factors, the fact that we have not confirmed the link between the CMPI and commodity prices may result from the omission of critical indicators. These may include not only different commodities but also measures of consumer sentiment such as data from Google Trends and similar sources.

Many of these limitations can be verified by repeating the research after more data has been collected. Clearly, over a longer period of time, the potential shortcomings are likely to become more conspicuous. Concurrently, it would be advantageous to broaden the scope of the research both geographically, by incorporating data from e-shops in other countries, and sectorally, by extending the focus towards indices such as the general CPI.

In the case of modelling constraints, it is certainly worth trying other types of models. We have considered a VAR model, but we consider it to be more suitable for cases where the system of variables interact significantly with each other. The price of electricity can certainly influence, for example, the price of copper mining, but the question is whether this is true of the price of electricity from the Prague exchange. Similarly, while it would make sense in theory that the price of manufacturing commodities is written into the price of building materials, we are on the fence about whether the reverse implication is also true.

To estimate the official index, which is published on monthly basis, it would certainly be worth testing the use of the MIDAS model that could estimate the monthly MIPI based on the weekly CMPI. Although Macias et al. (2022) are rather skeptical of MIDAS modeling of CPI (see Section 2.2.2), a check on the data for producer prices would be useful.

Similarly, because the data exhibit noticeable volatility, it would be interesting to explore a score-driven model on the data. The ARIMA model is based on the assumption of linear relationships, which may prove inadequate when analysing real-world data.

Therefore, it may be of interest to consider the potential of dynamic parameters that change over time, in order to capture the intrinsic dynamics of the data more accurately. For example, an investigation into the generalized autoregressive conditional ceteroskedasticity (GARCH) model, which accounts for volatility clustering, or the more comprehensive generalized autoregressive score (GAS) model may prove advantageous.

The GAS model, a broad framework that encompasses GARCH as a specific case, might be particularly beneficial due to the mentioned score-driven nature. This feature enables the model to adapt to new information effectively, ensuring that forecasts are updated and refined. Moreover, GAS models can establish a direct correlation between the shape of the conditional observation density and the dynamics of the parameters, which can enhance the robustness and precision of volatility estimates. The appropriateness of score-driven models in the context of macroeconomic data is shown, for example, in Li et al. (2017) in combination with the student distribution, or Blazsek et al. (2023) when cointegrating on US real GDP growth, US inflation rate, and effective federal funds rate data.

Chapter 5

Conclusion

The principal objective of this thesis was to investigate the potential enhancements in macroeconomic modeling facilitated by the integration of big data. To this end, we compiled a weekly price index of construction materials (CMPI) from web-scraped data collected over a period of almost three years.

Alternative to official indices

The first task was to verify the very possibility of measuring inflation of building materials through high-frequency data collected directly from retailers' websites by further extending the research described in the bachelor thesis (Štefl, 2022). The fact that the index prepared in this way can serve as an alternative proxy to the official price index of material inputs for construction works (MIPI) was confirmed during 144 weeks (11 quarters) of measurement. Among other things, the comparative scrutiny of these two indices revealed that the change in the Czech Statistical Office's (CZSO) weighting system at the onset of 2023 resulted in a permanent upward shift in the price level measured by the official index (see Figure 3).

The regular updating of the consumer basket is essential for maintaining the relevance of indices to contemporary economic realities. Nonetheless, our alternative metric challenges the comparability between the periods before and after modifications to the basket of goods of the official index, raising concerns about potential artificial overstatement of the inflation rate. This question is particularly pertinent given that the MIPI influences the general Consumer Price Index (CPI) through imputed rents. According to the CZSO, construction materials prices have increased by an average of 3.8% in 2023, primarily due to a significant price jump in January. In contrast, our CMPI shows a more modest annual increase of only 0.7%.

The weekly frequency of our data collection enabled us to more precisely identify the structural break following the covid-19 pandemic recovery and the immediate impacts of the Russian invasion of Ukraine. The progress of CMPI indicates that by 8 May 2022, approximately three months post the outbreak of the conflict, the building materials market had stabilized in terms of price level.

In this first part of the research, we successfully transfer the contributions of Macias et al. (2022), Powell et al. (2017), Breton et al. (2016), or Rigobon (2016) from consumer prices to the domain of producer prices.

Link between construction materials and commodity prices

The second objective of the thesis was to investigate whether prices of building materials (as measured by the CMPI) are interrelated with commodity prices. After a thorough examination of the effect of individual exogenous commodities, we distilled the price of electricity, diesel and copper as the variables with the largest effect.

However, using a pseudo-real time nowcasting experiment, we confirmed that none of our ARIMAX models can estimate the CMPI even one week ahead better than ARIMA without external variables (see Figure 8). The difference between the fit of the model over the entire period studied and the period after the structural break leads us to infer that in normal times, unaffected by pandemics or war conflicts, the price level of building materials is more dependent on customer sentiment than on prices of commodities required for their production. In the majority of model configurations, these exogenous variables exhibited a more pronounced influence on the training dataset including the period prior to the structural break compared to the dataset excluding it.

We deduce that incorporating commodity prices into the model may be rational during periods heavily impacted by external factors, where the influence of customers' demand diminishes. However, to substantiate a more definitive claim, further investigation over an extended time horizon would be necessary.

Thus, in this part of the research we reach similar conclusions as Durand & Blöndal (1988) for producer prices, namely that we do not find a clear short-run relationship between commodity prices and the corresponding price index. Hence, in any follow-up research, it would be worth focusing on longer-term relationships, similar to Boughton & Branson (1990), but this cannot be done after three years of measurement.

Nowcasting of the official price index

The third and most significant objective of our research was to use the one-month lead of the web-scraped CMPI to nowcast the official MIPI. The apparent similarity of the two indices suggested to the presence of cointegration, which was subsequently confirmed.

This finding enabled us to validate the positive impact of the CMPI on the precision of one-month ahead forecasts of the MIPI using the ARIMAX modeling framework over its benchmarks.

Over a testing period of approximately one and a half years, the average error, as quantified by the root mean square error (RMSE), decreased by about 25% compared to the best benchmark model that did not incorporate exogenous variables (see Table 10). This improvement is primarily attributed to the fact that whereas the mid-month MIPI is not published by the statistical office until around the middle of the following month, our web-scraped CMPI becomes available almost immediately after data collection, effectively providing a one-month lead time over the official figures.

Thus, we corroborate the findings of Modugno (2011) or Macias et al. (2022) that high-frequency data can enhance the accuracy of nowcasts for monthly price indices, albeit using quite different data sources in our analysis, firmly paving the way for novel techniques to improve nowcasting of producer prices

By and large, our findings confirm that high-frequency data obtained through web scraping holds significant value in the field of economics and can assist both policymakers and firms in making more informed decisions.

References

Ajmera R., Kook N., & Grilley J. **Impact of commodity price movements on CPI inflation**. *Monthly Labor Review*, 135, 29. 2012. Link: https://www.bls.gov/opub/mlr/2012/04/art3full.pdf

Angelini E., Bańbura M., & Rünstler G. **Estimating and forecasting the euro area monthly national accounts from a dynamic factor model.** *ECB Working Paper No. 953. European Central Bank*. 2008. Link: https://www.econstor.eu/bitstream/10419/153387/1/ecbwp0953.pdf

Bertolotto M., & Aparicio D. **Forecasting inflation with online prices.** *International Journal of Forecasting,* 36(2), April–June 2020, 232-247. 2020. Link: https://doi.org/10.1016/j.ijforecast.2019.04.018

Atkeson A., & Ohanian L. **Are Phillips Curves Useful for Forecasting Inflation?** *Federal Reserve Bank of Minneapolis, Quarterly Review.* 2001. Link: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=63940a2734d944b28c6ae8ad3ae69ea9ab3b317f

Beckerman W., & Jenkinson T. **What Stopped the Inflation? Unemployment or Commodity Prices?** *Economic Journal*, Vol. 96 (March 1986), pp. 39-54. 1986. Link: https://doi.org/10.2307/2233425

Beeson J. **Web Scraped Data: Extreme price changes.** *Office for National Statistics*. 2015. Link: https://www.ons.gov.uk/file?uri=/aboutus/whatwedo/programmesandprojects/theonsbigdataproject/webscrapeddataextremepricechanges_tcm77-415882.pdf

Bhardwaj H., Flower T., Lee P., & Mayhew M. **Research indices using web scraped price data: August 2017 update**. *Office for National Statistics*. 2017, 1.9.2017. Link: https://www.ons.gov.uk/economy/inflationandpriceindices/articles/researchindicesusingwebscrapedpricedata/august2017update

Bird D., Breton R., Payne C., & Restieaux A. **Initial report on experiences with scanner data in ONS.** *Office for National Statistics*. 2014, 16.12.2014. Link: https://webarchive.nationalarchives.gov.uk/20160106041638/http:/www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/initial-report-on-experiences-with-scanner-data-in-ons.pdf

Blazsek S., Escribano A., & Licht A. **Co-integration with score-driven models: an application to US real GDP growth, US inflation rate, and effective federal funds rate**. *Macroeconomic Dynamics*, Cambridge University Press, vol. 27(1). 2023. Link: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/A1356C9CA70493D3A43758FE8A3300BB/S1365100521000365a.pdf/cointegration_with_scoredriven_models_an_application_to_us_real_gdp_growth_us_inflation_rate_and_effective_federal_funds_rate.pdf

Bosworth B. P., & Lawrence Z. R. **Commodity Prices and the New Inflation**. *The Brookings Institution*. Washington. 1982. Link: https://www.cabidigitallibrary.org/doi/full/10.5555/19831803743

Boughton M. J., & Branson H. W. **The Use of Commodity Prices to Forecast Inflation**. *International Monetary Fund*. 1990. Link: https://www.elibrary.imf.org/display/book/9781451943818/ch01.xml

Breton R., Clews G., Metcalfe E., Milliken N., Payne C., Winton J., & Woods A. **Research indices using web scraped data**. *Office for National Statistics*. 2015, 23.10.2015. Link: https://webarchive.nationalarchives.gov.uk/ukgwa/20160106041631/http://www.ons.gov.uk/ons/rel/cpi/consumer-price-indices/research-indices-using-web-scraped-price-data/index.html

Breton R., Flower T., Mayhew M., Metcalfe E., Milliken N., Payne C., Smith T., Winton J., & Woods A. **Research indices using web scraped data: May 2016 update**. *Office for National Statistics*. 2016, 23.5.2016. Link: https://backup.ons.gov.uk/wp-content/uploads/sites/3/2016/06/researchindicesusingwebscrapeddatamay2016.pdf

Canova F. **Methods for Applied Macroeconomic Research.** *Princeton University Press*. 2007. ISBN: 978-0-691-11504-7

Cavallo A. **Online and official price indexes: Measuring Argentina's inflation**. *Journal of Monetary Economics*. *Massachusetts Institute of Technology*. 2012, 9.10.2012. Link: http://dx.doi.org/10.1016/j.jmoneco.2012.10.002

Cavallo A. **Scraped data and sticky prices**. *National bureau of economic research*. 2015. Link: https://www.nber.org/system/files/working_papers/w21490/w21490.pdf

Cavallo A., & Rigobon R. **The Billion Prices Project: Using Online Prices for Measurement and Research**. *Journal of Economic Perspectives*, 30 (2). 2016. Link: https://www.aeaweb.org/articles?id=10.1257/jep.30.2.151

Cavallo A. **Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers**. *American Economic Review,* 107 (1), 283–303. 2017. Link: https://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.20160542

Consumer Price Index: Calculation. *Bureau of Labor Statistics*. 2023. Link: https://www.bls.gov/opub/hom/cpi/calculation.htm#:~:text=Cell%2Drelative%20imputation&text=The%20price%20change%20between%20the,for%20that%20ELI%20and%20PSU.

Durand M., & Blöndal S. **Are Commodity Prices Leading Indicators of OECD Prices?** *OECD Working Paper,* No. 49, OECD Department of Economics and Statistics, 1988. Link: https://web-archive.oecd.org/2012-06-15/147978-2348535.pdf

Einav L., & Levin J. **Economics in the age of big data**. *Science*, 346 (6210). 2014, 7.11.2014. Link: https://www.science.org/doi/10.1126/science.1243089

Faust J., Wright H. J., **Chapter 1 - Forecasting Inflation**. *Handbook of Economic Forecasting,* vol 2, part A. 2-56. 2013. ISSN 1574-0706, ISBN 9780444536839. Link: https://doi.org/10.1016/B978-0-444-53683-9.00001-3.

Feenstra C. R. **New Product Varieties and the Measurement of International Prices**. *American Economic Review*, vol 84. March 1994, 157-177. 1994. Link: https://www.science.org/doi/10.1126/science.1243089

Fisher D. M. J., Liu T. C., & Zhou R. **When can we forecast inflation?** *Economic Perspectives*. 2002. Link: https://www.chicagofed.org/~/media/publications/economic-perspectives/2002/1qepart4-pdf.pdf

Franta M., Sútoris I. **Dynamics of Czech Inflation: The Role of the Trend and the Cycle**. *CNB Working Paper*, July 2020. 2020 Link: https://www.cnb.cz/export/sites/cnb/en/economic-research/.galleries/research_publications/cnb_wp/cnbwp_2020_01.pdf

Goolsbee A. D., & Klenow J. P. **Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce**. *American Economic Association*. 2018. Link: https://www.aeaweb.org/articles?id=10.1257/pandp.20181038

Giannone D., Reichlin L., & Small D. **Nowcasting: The real-time informational content of macroeconomic data**. Journal of Monetary Economics, 55(4), 665–676. 2007. Link: http://www.indexmeasures.ca/dc2008/papers/Nowcasting.pdf

Groen J., Paap R., & Ravazzolo F. **Real-Time Inflation Forecasting in a Changing World**. *Journal of Business & Economic Statistics*. 2013, 31 (1), 29-44. Link: https://doi.org/10.1080/07350015.2012.727718

Hayes B. **Research indices using web scraped price data: clothing data** *Office for National Statistics*. 2017. Link: https://www.ons.gov.uk/economy/inflationandpriceindices/articles/researchindicesusingwebscrapedpricedata/clothingdata

Hobijn B. **Commodity price movements and PCE inflation**. *Current Issues in Economics and Finance*, 2008, 14(8). Federal Reserve Bank of New York. 2008. Link: https://www.newyorkfed.org/research/economists/medialibrary/media/research/current_issues/ci14-8.pdf

How did the CNB's two-week repo rate develop?. *Czech National Bank*. Prague, 2024. Link: https://www.cnb.cz/cs/casto-kladene-dotazy/Jak-se-vyvijela-dvoutydenni-repo-sazba-CNB/

Inflation, Inflation Rate: Methodology. *Czech Statistical Office*. Prague, 2024. Link: https://www.czso.cz/csu/czso/what_is_it_inflation_resp_inflation_rate

Inflation Rate. *Czech Statistical Office*. 2024. Link: https://www.czso.cz/csu/czso/inflation-rate

Li G., Li Z., Wu D. D., Zhou J. **Forecasting tourism demand with cross-country monthly data.** *International Journal of Forecasting*, vol 33. January 2017, 43-52. 2017. Link: https://doi.org/10.1016/j.ijforecast.2016.11.007

Konny G. C., Williams B. K., & Friedman D. M. **Big Data in the U.S. Consumer Price Index: Experiences & Plans**. *U.S. Bureau of Labor Statistic*s. 2019. Link: https://www.brookings.edu/wp-content/uploads/2019/02/Big-Data-in-the-U.S.-Consumer-Price-Index.pdf

Knotek S. E., & Zaman S. **Nowcasting U.S. Headline and Core Inflation.** *Federal Reserve Bank of Cleveland, working paper no. 14-03R*. 2015. Link: https://www.clevelandfed.org/-/media/project/clevelandfedtenant/clevelandfedsite/publications/working-papers/2015/wp-1403r-nowcasting-us-headline-and-core-inflation-pdf.pdf

Lünnemann P., & Wintr L. **Are internet prices sticky?** *ECB Working Paper No. 645. European Central Bank*. 2006. Link: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp645.pdf

Macias F., & Stelmasiak D. **Food inflation nowcasting with web scraped data**. *Narodowy Bank Polski: Education & Publishing Department*. 2019. Link: https://static.nbp.pl/publikacje/materialy-i-studia/302_en.pdf

Macias F., Stelmasiak D., & Szafranek K. **Nowcasting food inflation with a massive amount of online prices**. *Narodowy Bank Polski: Education & Publishing Department*. 2022, 28.3.2022. Link: https://doi.org/10.1016/j.ijforecast.2022.02.007

Mankiw, G. **Principles of Economics**. 1997. ISBN 0-030-27087-1

Mayhew M. **Imputing Web Scraped Prices**. *Office for National Statistics*. 2016, 23.5.2016. Link: https://www.ons.gov.uk/file?uri=/economy/inflationandpriceindices/methodologies/imputingwebscrapedprices/imputingwebscrapedpricesmayhewm.pdf

Melser D. **Constructing High Frequency Price Indexes Using Scanner Data**. *School of Finance & Economics. University of Technology Sydney*. 2011, 27.4.2011, . Link: https://www.ottawagroup.org/ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/b1ab2e631d34d9bbca2578a7007fa493/$FILE/2011%2012th%20meeting%20-Daniel%20Melser%20(School%20of%20Finance%20&%20Economics,%20University%20of%20Technology,%20Sydney)%20Constructing%20High%20Frequency%20Price%20Indexes%20Using%20Scanner%20Data.pdf

Metcalfe E., Flower T., Lewis T., Mayhew M., & Rowland E. **Research indices using web scraped price data: clustering large datasets into price indices (CLIP)**. *Office for National Statistics*. 2016, 30.11.2016. Link: https://www.ons.gov.uk/economy/inflationandpriceindices/articles/researchindicesusingwebscrapedpricedata/clusteringlargedatasetsintopriceindicesclip

Methodical notes – Producer price indices. *Czech Statistical Office*. Prague, 2022. Link: https://www.czso.cz/documents/10180/142756936/01104121q4m1en.pdf/0b78eee2-f023-4e9d-954e-38294e3424dc?version=1.1

Meyler A., Kenny G., & Quinn T. **Forecasting Irish inflation using ARIMA models**. *Central Bank and Financial Services Authority of Ireland.* 1998. Link: https://mpra.ub.uni-muenchen.de/11359/1/MPRA_paper_11359.pdf

Modugno M. **Nowcasting inflation using high frequency data**. *ECB Working Paper No. 1324, European Central Bank.* 2011. Link: https://www.econstor.eu/bitstream/10419/153758/1/ecbwp1324.pdf

Monteforte L., & Moretti G. **Real time forecasts of inflation: the role of financial variables.** *LLEE Working Document No. 81, June 2009. Luiss Lab of European Economics.* 2009. Link: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=fad9d0020fd12e8076eea1e4ad01b03f64f5df9e

Mrázek, J. **Consumer Price Indices: User's methodological manual**. *Czech Statistical Office*. Prague, 2021, 8.9.2021. Link: https://www.czso.cz/documents/10180/26822363/manual_isc_2021en.pdf/953a150f-fad5-47db-960d-65a19d1de101?version=1.6

Payne C. **Analysis of product turnover in web scraped clothing data, and its impact on methods for compiling price indices.** *Office for National Statistics*. 2017. Link: https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/currentmethodologyarticles/analysisofproductturnoverinwebscrapedclothingdataanditsimpactonmethodsforcompilingpriceindices

Powell B., Nason G., Elliot G., Mayhew M., Davies J., & Winton J. **Tracking and modelling prices using web-scraped price microdata: towards automated daily consumer price index forecasting**. *Journal of the Royal Statistical Society*, series A, (Statistics in Society), 181(3). 2017. Link: https://rss.onlinelibrary.wiley.com/doi/10.1111/rssa.12314

Quarterly price indices of construction works and constructions. *Czech Statistical Office*. 2022. Link: https://www.czso.cz/documents/10180/37695677/mipccr100419_4_aj.pdf/9c2ff12e-707b-4b7c-998e-1067f78d15ae?version=1.1

Štefl, Josef. **High Frequency Price Index of Construction Materials**. Bachelor's thesis. *Charles University, Faculty of Social Sciences, Institute of Economic Studies*. Prague, 2022. Advisor: Mgr. Petr Polák, M.Sc., Ph.D., M.Sc. Link: http://hdl.handle.net/20.500.11956/175640

Tang X., Wang L., Cheng J., & Chen J. **Forecasting model based on information-granulated GA-SVR and ARIMA for producer price index**. *arXiv preprint*, arXiv:1903.12012, Hainan University. 2019. Link: https://arxiv.org/pdf/1903.12012

Wei, Shang-Jin. **Understanding a Divergence between CPI and PPI.** *National Bureau of Economic Research.* 2019. Link: https://www.nber.org/system/files/working_papers/w24319/w24319.pdf

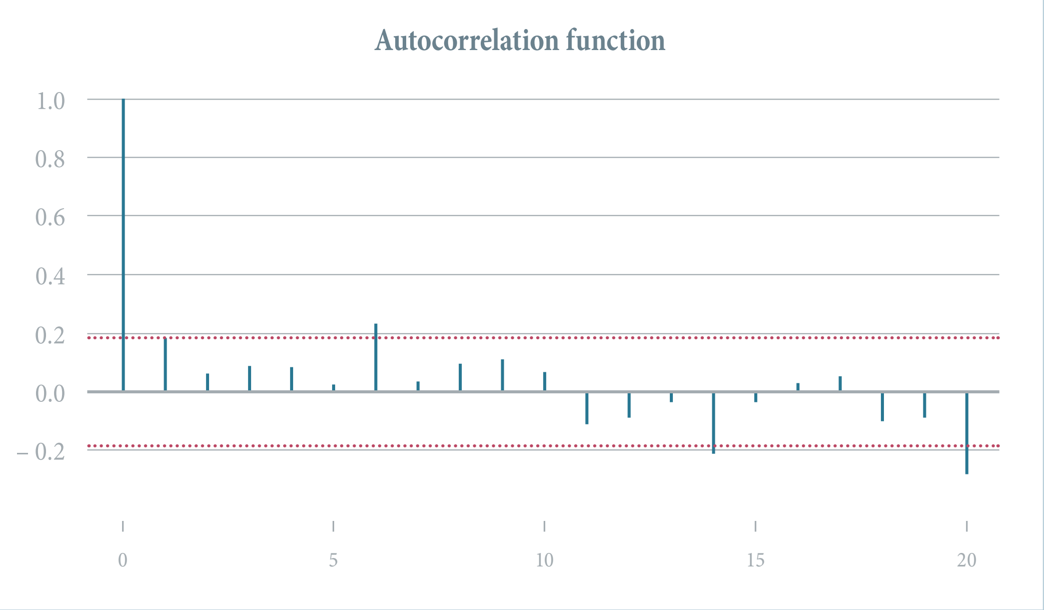
Zhao B., **Web Scraping**, *Encyclopedia of Big Data*. Springer International Publishing. 2017. Link: https://www.researchgate.net/publication/317177787_Web_Scraping

Appendix A

Autocorrelation

Figure 12: ACF and PACF of week-on-week CMPI

October 2021 – June 2024, weekly frequency (CMPI WoW)



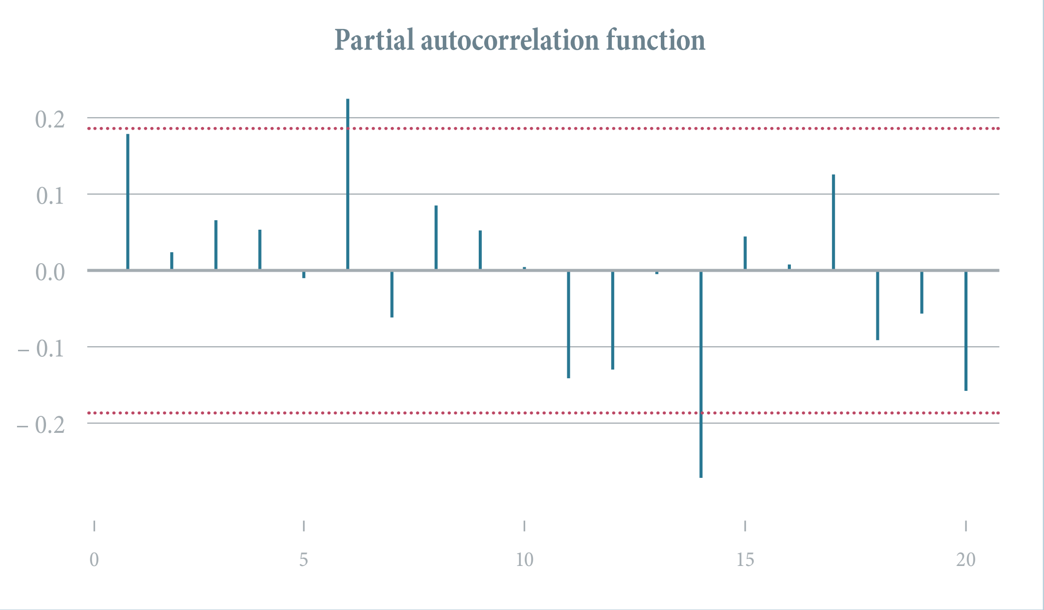
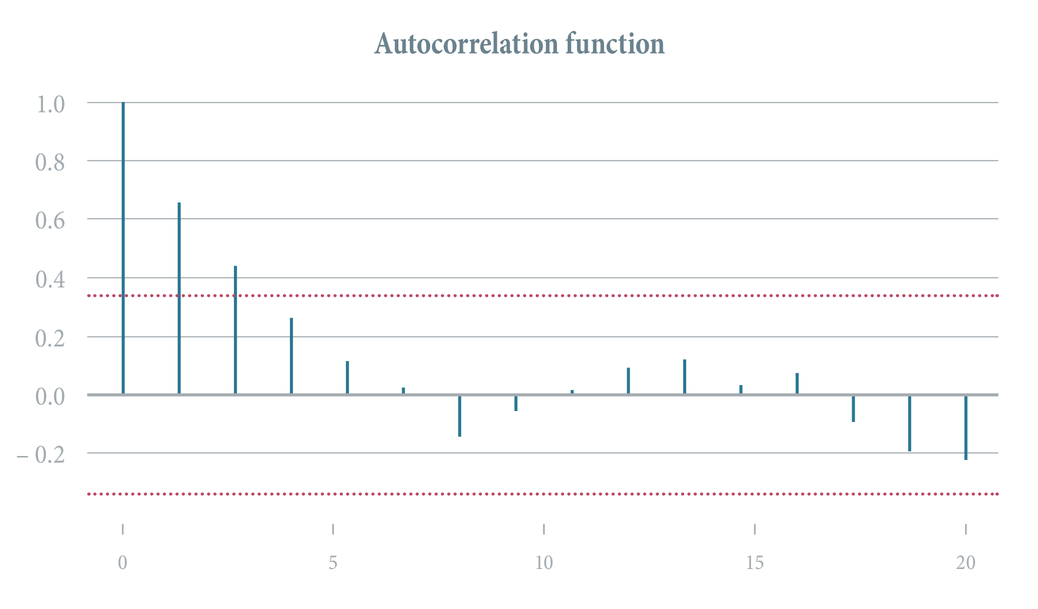


Figure 12 shows the results of autocorrelation and partial autocorrelation functions of week-on--week changes in CMPI. The plot suggests that the strong AR pattern is not present, which is also true for long tails or gradual declines that typically indicate MA processes.

Figure 13: ACF and PACF of month-on-month MIPI

October 2021 – June 2024, monthly frequency (MIPI MoM)



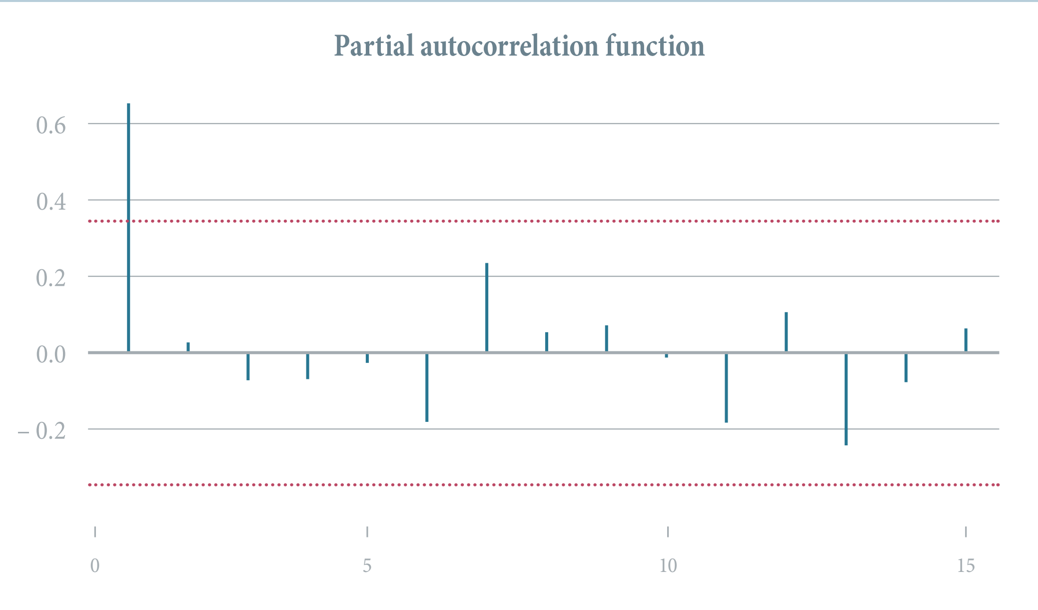


Figure 13 shows the results of autocorrelation and partial autocorrelation functions of month-on--month changes in the official MIPI. The significant spike at lag 1 in both ACF and PACF suggests that an AR(1) component is appropriate for the model, indicating an ARIMA model with p=1. The gradual decline in lagged effects also implies the influence of the MA component.

Appendix B

Post-break dependencies

Table 11: Results of Granger causality test on post-break period

May 2022 – June 2024, weekly frequency (CMPI WoW), MA – Moving Average, W – Week, M – Month

Commodity	Smoothing	Basis	Lag	*p*-value
Diesel	MA (week)	WoW	3 months	0,0064
Iron Ore		WoW	1 week	0,0088
Electricity	MA (week)	WoW	1 month	0,0250
Natural Gas	MA (week)	WoW	1 month	0,0288
Petrol	MA (week)	WoW	3 months	0,0294
Aluminum	MA (month)	WoW	3 months	0,0772
Lumber	MA (month)	WoW	3 months	0,1336
Crude Oil_{WTI}	MA (month)	WoW	2 months	0,1474
Crude Oil_{Brent}	MA (month)	WoW	2 months	0,1530
Copper	MA (week)	WoW	3 months	0,2056
Steel		WoW		0,2065

Table 11 summarizes the p-values of the Granger causality test for the 11 variables in relation to the CMPI on restricted period after the structural break. For each commodity, only the best-fit representation is reported. While only *Lumber* and *Copper* prices are rejected at the 5% significance level in the full sample (see Table 5), in the standard period, i.e. not affected by post-pandemic recovery or war conflict, it appears that in addition to the aforementioned commodities, *Aluminum*, *Crude Oil*, and *Steel* also do not Granger-cause the CMPI. *Diesel*, on the other hand, retains almost the same significance for explaining CMPI in both datasamples.

Appendix C

Consumer basket

CB ID	CMPI ID	Item name	Volume	Weight
1	3	Lišta rohová R2 smrk nastavený 35×19×2000 mm, 25 ks/bal.	74.7 m	0.05%
2	3	Lišta podlahová P5 smrk nastavený 27×9×2000 mm, 25 ks/bal	267.9 m	0.19%
3	6	Deska cementotřísková CETRIS BASIC 12×1250×3350 mm	64.9 m^2	0.29%
4	7	Deska KRONOSPAN OSB 3 15×2500x1250 mm	42 m^2	0.20%
5	9	Deska polykarbonátová trapézová TOPLIGHT T 76/18 hladká 2UV čirá 1,06×4 m	24 m^2	0.06%
6	11	Páska akustická na profil 75 DK Mont 30 m	60 m	0.01%
7	12	Deska sádrokartonová Knauf WHITE GKB 12,5×1250×2000 mm	246.3 m^2	0.22%
8	12	Deska sádrokartonová akustická Knauf Blue Akustik 12,5×1250×2000 mm	15.8 m^2	0.02%
9	14	Masivní konstrukční dřevo KVH Si, smrk	6.4 m^3	3.03%
10	14	Hranol smrkový s povrchem po řezu	1 m^3	0.20%
11	16	Profily z masivního dřeva KVH NSi 100x100x5000 mm (44 ks/pak.)	0.2 m^3	0.09%
12	16	Masivní konstrukční dřevo KVH Si, smrk	1.5 m^3	0.71%
13	16	Profily z masivního dřeva KVH Nsi 40x60x5000 mm	0.2 m^3	0.09%
14	16	Profily z masivního dřeva KVH NSi 40x60x4000 mm (180ks/pak.)	0.5 m^3	0.18%
15	16	Profily z masivního dřeva KVH NSi 40x60x4000 mm (180ks/pak.)	228 m	0.24%
16	16	Masivní konstrukční dřevo KVH Si, smrk	0.5 m^3	0.23%
17	16	Hranol pod terasy AU-MEX exotické dřevo 45×70×3 970 mm	80.4 m	0.18%
18	17	Palubka obkladová SECA A/B klasik smrk 19×146×4000 mm (5 ks/bal)	148 m^2	0.23%
19	17	Palubka obkladová SECA A/B klasik smrk 14×121×4000 mm (8 ks/bal)	134.9 m^2	0.79%
20	17	Palubka obkladová SECA A/B klasik smrk 24×146×4000 mm (4 ks/bal)	41 m^2	0.31%
21	18	Bednění prkenné 24/22×min. 80 mm délka 5 m	0.8 m^3	0.13%
22	18	Prkno terasové AU-MEX Massaranduba 25×145×3660 mm	37 m^2	0.17%
23	19	Střešní lať ze smrkového dřeva 30x50/5000 mm impregnovaná	1.2 m^3	0.23%
24	19	Střešní lať ze smrkového dřeva 40x60/5000 mm impregnovaná	1.1 m^3	0.45%
25	22	Profil zakládací Levelys 1 500 mm	25.9 m	0.22%
26	22	Hliníkový roh s tkaninou Likov Vertex 100x2000 mm	32.2 m	0.01%
27	22	Lišta hliníková L TopStone 15 mm, 2,5 m	17.5 m	0.02%
28	22	Lišta okenní začišťovací PS-VH 0,6x2400 mm	21 m	0.01%
29	22	Lišta okenní začišťovací LS-EKO Vertex 2400 mm	56.7 m	0.04%
30	23	Talířová hmoždinka TK-PSV 8×140 mm (20 ks/bal)	524 pcs	0.10%
31	23	Páska lepicí DEKTAPE FASSADE šířka 60 mm délka 25 m	107.1 m	0.04%
32	24	Tkanina VERTEX R85 110 g/m2 (50 m2/bal.)	526.2 m^2	0.29%
33	26	Hydroizolace Sikalastic 612, šedá, 5 l	113.5 kg	0.09%
34	26	Lak izolační Renolak ALN 9 kg/bal.	338.4 m^2	0.11%
35	26	GUMOASFALT SA 12 (5kg/bal.)	110 kg	0.06%
36	26	Hydroizolace Ceresit CL 50 Express 2-K 12,5 kg	21 kg	0.03%
37	26	Stěrka hydroizolační Mapei Mapegum WPS 20 kg	28 kg	0.05%
38	27	Samolepicí asfaltový pás z SBS modifikovaného asfaltu TOPDEK AL BARRIER	67.9 m^2	0.13%
39	27	Oxidovaný asfaltový pás se skleněnou tkaninou DEKGLASS G200 S40 (role/7,5m2)	67.9 m^2	0.09%
40	27	Oxidovaný asfaltový pás DEK R13 (role/20m2)	224 m^2	0.22%

CB ID	CMPI ID	Item name	Volume	Weight
41	27	Asfaltový pás GLASTEK AL 40 MINERAL (role/7,5 m2)	147.8 m^2	0.31%
42	27	Oxidovaný asfaltový pás DEK R13 (role/20m2)	140.7 m^2	0.07%
43	27	Oxidovaný pás DEKBIT AL S40 (role/7,5 m2)	154 m^2	0.22%
44	27	Samolepicí asfaltový pás GLASTEK 30 STICKER PLUS KVK (role/10 m2)	164.5 m^2	0.34%
45	27	Asfaltový pás parotěsnicí samolepicí TOPDEK AL BARRIER (7,5 m2/role)	156.1 m^2	0.31%
46	28	Geotextilie FILTEK 400 g/m2, šíře 2 m	57.2 m^2	0.05%
47	28	Geotextilie GEOTEK Z 500 g/m2, šíře 2 m	14.3 m^2	0.01%
48	28	Geotextilie GEOTEK Z 300 g/m2, šíře 2 m	140 m^2	0.06%
49	29	Izolace spodní stavby z PVC-P ALKORPLAN 35034 2,0 mm, šíře 2,15 m	627.9 m^2	1.07%
50	29	Fólie hydroizolační MAPEPLAN T WT světle zelená 1,5 mm šíře 2,1 m	70 m^2	0.33%
51	30	DEKDREN L60 GARDEN profilovaná (nopová) fólie s perforací	20.1 m^2	0.10%
52	30	Fólie nopová DEKDREN G8 výška nopu 8 mm šířka 2,0 m (40 m2/bal.)	70 m^2	0.12%
53	31	Parozábrana DEKFOL N 140 STANDARD 75 m2/bal.	336 m^2	0.14%
54	31	Parozábrana JUTAFOL N AL 170 SPECIÁL A.P. (75m2/bal.)	124.9 m^2	0.13%
55	31	Fólie separační, DEKSEPAR	59 m^2	0.01%
56	33	Oboustranná spojovací páska JUTAFOL SP1 (15mm x 45m) na parozábrany	83.3 m	0.01%
57	33	Páska PVC těsnicí Mapei Mapeband PE 120 50 m	66.5 m	0.06%
58	33	Páska těsnicí BRAMAC pod kontralatě šířka 50 mm	226.5 m	0.10%
59	34	Pěnová páska TP600 illmod 600, šedá, 13-24x25mm, délka 5,2m	14 m	0.05%
60	36	Tepelná izolace Fibran XPS L 300 kPa 50 mm (6 m2/bal.)	9.8 m^2	0.03%
61	36	Extrudovaný polystyren FIBRAN 300-L 100 mm (1250x600 mm)	14 m^2	0.09%
62	37	Tepelná izolace K-FLEX ST 140/50 mm	105 m	0.04%
63	37	Pouzdro izolační 65 Al 108/80 mm	37.8 m	0.18%
64	37	Tepelná izolace TUBEX STANDARD 48/10 mm, délka 2m	56 m	0.02%
65	38	Tepelná izolace ISOVER ORSIK 80 mm (4,5 m2/bal.)	65.6 m^2	0.10%
66	38	Tepelná izolace DEKWOOL G 039r role 100 mm (11,76 m2/bal)	147.7 m^2	0.27%
67	38	Tepelná izolace DEKWOOL G 039r 120 mm (9,84 m2/bal.)	93.9 m^2	0.30%
68	38	Tepelná izolace DEKWOOL DW r plate 60 mm (15,63 m2/bal)	44.8 m^2	0.03%
69	38	Tepelná izolace ISOVER DOMO PLUS 100 mm (10,08 m2/bal.)	478.8 m^2	0.68%
70	38	Izolace kročejová ISOVER T-P 30 mm (5,04 m2/bal)	81.1 m^2	0.21%
71	38	Izolace kročejová ISOVER T-N 40 mm (4,32 m2/bal)	1.6 m^2	0.01%
72	38	Tepelná izolace ISOVER UNI desky 50 mm (7,2 m2/bal)	12.2 m^2	0.01%
73	38	Izolace kročejová ISOVER T-P 20 mm (7,2 m2/bal)	166.9 m^2	0.14%
74	38	Tepelná izolace ISOVER UNI desky 100 mm (3,6 m2/bal)	23.5 m^2	0.05%
75	38	Tepelná izolace DEKWOOL DW r plate 100 mm (9,38 m2/bal)	10.3 m^2	0.14%
76	38	Tepelná izolace DEKWOOL DW r plate 140 mm (6,25 m2/bal)	264.5 m^2	0.42%
77	38	Atikový klín z minerální vaty 50×50 mm	32.2 m	0.15%
78	38	Tepelná izolace DEKWOOL DW r plate 80 mm (11,72 m2/bal)	23.6 m^2	0.13%
79	40	Tepelná izolace Dekperimeter SD 150 mm (2,25 m2/bal.)	58.6 m^2	0.27%
80	40	Tepelná izolace Isover EPS 70 F 160 mm (1,5 m2/bal.)	21 m^2	0.09%
81	40	Polystyren EPS 150 50 mm (500×1000 mm)	12.1 m^2	0.08%
82	40	Tepelná izolace Isover EPS 100 50 mm (5 m2/bal.)	21.8 m^2	0.03%
83	40	Tepelná izolace Isover EPS 150 100 mm (2,5 m2/bal.)	30.5 m^2	0.12%
84	40	Tepelná izolace Isover EPS 100 200 mm (1 m2/bal.)	79.9 m^2	0.50%
85	40	Fasádní polystyren DEK EPS 70F 60 mm (1000×500 mm)	9.2 m^2	0.01%
86	40	Tepelná izolace Bachl EPS 100 20 mm (12,5 m2/bal.)	14 m^2	0.01%
87	40	Tepelná izolace P-Systems EPS 200 110 mm (2,5 m2/bal.)	128.8 m^2	0.64%

CB ID	CMPI ID	Item name	Volume	Weight
88	40	Tepelná izolace Isover EPS GreyWall Plus 160 mm (1,5 m2/bal.)	14 m^2	0.11%
89	40	Tepelná izolace P-Systems EPS 70 110 mm (2,5 m2/bal.)	112 m^2	0.32%
90	42	Deska z kamenné vlny ISOVER FIREPROTECT 150, tloušťka 20 mm	32.7 m^2	0.07%
91	42	Deska z kamenné vlny s hliníkovým polepem ISOVER ORSTECH 65H, 40 mm	233.5 m^2	0.47%
92	42	Tepelná izolace TOPDEK SKY 140 mm (5,952 m2/bal.)	42.5 m^2	0.65%
93	42	Deska sanační IsoAir 50 mm (950×950 mm)	2.3 m^2	0.03%
94	42	Deska sanační IsoAir 80 mm (950×950 mm)	23.6 m^2	0.48%
95	42	Izolace podlahová Fenix F-Board 6 mm 7,2 m2	11.8 m^2	0.07%
96	42	Deska systémová Uponor SICCUS (12,5 m2/bal.)	9.1 m^2	0.03%
97	42	Deska systémová s izolací Gabotherm GTF-TAC 30-2 (role 10 m2)	14 m^2	0.03%
98	42	Tepelná izolace Baumit openReflect 220 mm (1 m2/bal.)	4.2 m^2	0.30%
99	43	Termoizolační pás Mirel Vratimov MIRELON C3 bílý tl. 3 mm, 110 cm x 100 m	11.3 m^2	0.00%
100	43	Termoizolační pás Mirel Vratimov MIRELON C3 bílý tl. 2 mm, 110 cm x 100 m	97.2 m^2	0.02%
101	49	Lišta ukončovací čtvercová Acara nerez 10 mm	83.8 m	0.19%
102	49	Lišta schodová protiskluzová Z Acara hliník 10 mm	42 m	0.04%
103	51	Obklad KAI UTOPIA 25×40 cm beige KAI.5970	73 m^2	0.27%
104	51	Obklad MINERALS 20×40 cm šedá WADMB436.1	65 m^2	0.22%
105	51	Obklad Rako Extra 20×40 cm tmavě šedá WADMB724	55 m^2	0.25%
106	51	Sokl Rako Faro 60×7,2 cm šedo-bílá DSASP719	25.8 m	0.13%
107	51	Obklad Kanjiža HABITAT 25×50 cm noce KJ-HA-NOC25	26.5 m^2	0.14%
108	51	Dekor Rako QUARZIT 30×60 cm šedá DDTSE737	31 m	0.39%
109	54	Dlažba Rako QUARZIT 60×60 cm tmavě šedá DAR63738	26 m^2	0.31%
110	54	Dlažba KAI SUBWAY 60×120 cm light grey KAI.9923	42 m^2	0.36%
111	54	Dlažba Rako Block 60×60 cm světle šedá DAK63780	28 m^2	0.25%
112	54	Dlažba Rako Alba 30×60 cm hnědo-šedá DAPSE732	35.1 m^2	0.34%
113	55	Podlaha dřevěná Feel Wood dub markant 137×2 053×21 mm	103.4 m^2	3.01%
114	55	Podlaha dřevěná EkoWood Classic bílá, 192×2 150×13,5 mm	125 m^2	1.13%
115	60	Lišta dělící T Acara eloxovaný hliník	75 m	0.09%
116	60	Lišta soklová Feel Wood dub 2 bm	84 m	0.18%
117	62	Podlaha vinylová lepená Home sahara oak brown	19.7 m^2	0.13%
118	62	Podlaha vinylová zámková SPC Home sahara oak brown	32 m^2	0.39%
119	64	Podsyp vyrovnávací Fermacell 50 l	90 kg	0.02%
120	67	Dveře kašírované plné levé bílé oblá hrana voština – 60 cm	3 pcs	0.04%
121	67	Dveře interiérové plné Polskone ARCO pravé 700 mm dub halif	2 pcs	0.15%
122	67	Dveře interiérové Polskone INTER AMBER prosklené, pravé dub sonoma 800 mm	1 pcs	0.07%
123	67	Dveře interiérové plné Polskone INTER AMBER levé 700 mm dub delano	5 pcs	0.18%
124	67	Dveře plné hladké Masonite DTD CPL bílé premium pravé 800 mm	2 pcs	0.11%
125	67	Dveře plné hladké Masonite DTD CPL bílé premium pravé 900 mm	1 pcs	0.06%
126	67	Dveře kašírované plné pravé bílé oblá hrana voština – 80 cm	3 pcs	0.04%
127	68	Dveře ocelové plné zateplené levé šířka 800 mm bílé	1 pcs	0.08%
128	72	Okno plastové VIVA LINE bílé/antracit levé 1 200 × 1 200 mm	3 pcs	0.21%
129	72	ACO Okno plastové 80x40 cm bílé	1 pcs	0.03%
130	72	ACO Markant okno plastové SK hnědé 40x40 cm	1 pcs	0.04%
131	72	Okno plastové VIVA LINE bílé/antracit levé 900 × 1 200 mm	5 pcs	0.31%
132	72	ACO Markant okno plastové SK bílé 100x60 cm	1 pcs	0.04%
133	72	ACO Markant okno plastové bílé 60x60 cm	2 pcs	0.06%
134	74	Okno profilované BRAMAC Classic Luminex cihlově červená	1 pcs	0.06%

CB ID	CMPI ID	Item name	Volume	Weight
135	74	Okno BRAMAC Luminex UNI 705×765 mm cihlově červená	4 pcs	0.28%
136	76	Světlovod pro plochou střechu SUNIZER ROUND průměr 430 mm	2 pcs	0.66%
137	76	Světlovod pro šikmou střechu SUNIZER ROUND průměr 530 mm	1 pcs	0.44%
138	77	Dvířka vanová lakovaná LTM 200×200 mm	3 pcs	0.02%
139	81	Schody půdní Roto ESCA 11 ISO RC 1 200×600 mm dřevěné	2 pcs	0.17%
140	89	Pouzdro Eclisse Standard KIT 600/1970 mm do SDK tl. 100–125 mm	2 pcs	0.20%
141	91	Zárubeň ocelová DEK YH 100 DV levá 700 mm	5 pcs	0.11%
142	91	Zárubeň ocelová DEK YH 150 DV pravá 800 mm	2 pcs	0.06%
143	91	Zárubeň ocelová DEK YH 150 DV pravá 900 mm	1 pcs	0.03%
144	91	Zárubeň ocelová DEK YH 75 DV pravá 900 mm	2 pcs	0.05%
145	91	Zárubeň ocelová DEK YH DV 150 pravé, šířka 900 mm	3 pcs	0.09%
146	94	Stohovatelný terč pod dlažbu 157/15 mm	25 pcs	0.00%
147	94	Průchodka rovnou střechou DN 125	2 pcs	0.02%
148	94	Vtok střešní TPE Dutral průměr 125 mm, délka 600 mm	1 pcs	0.01%
149	94	Perforovaný ochranný koš TOPWET TWOK v100	1 pcs	0.01%
150	94	Odvětrárání kanalizace TOPWET TWOP 50 BIT	2 pcs	0.03%
151	94	Sněhový zachytávač RS250S SP25 RAL3011 0,50mm 2bm	28 pcs	0.26%
152	94	Sněhový zachytávač RS250S SP35 2R17A 0,50mm 2bm	40 pcs	0.49%
153	94	Stohovatelný terč pod dlažbu 157/15 mm	300 pcs	0.03%
154	96	Šindel asfaltový IKO Victorian Plus 01 Černá 2,0 m2	130 m^2	0.51%
155	98	Taška střešní Röben Piemont engoba hnědá	200 m^2	3.70%
156	98	Hřebenáč napojovací tvar Y Röben tobago glazura	50 pcs	1.59%
157	103	Stříška vchodová JAP RAIN 1800×1000 mm	2 pcs	0.58%
158	106	Omítka minerální Baumit SilikatTop škrábaná 2 mm 25 kg	223 kg	0.20%
159	106	Omítka pastovitá weberpas silikon zrnitá 3 mm BI00 25 kg	23.6 m^2	0.08%
160	106	Akrylátová rýhovaná omítka PCI Multiputz RA 1,5mm, bílá, 25kg	315 kg	0.24%
161	106	Omítka pastovitá weberpas ExtraClean Active zrnitá 1 mm bílá	572.2 m^2	1.12%
162	107	Omítka jádrová Baumit Manu 2 vápenocementová 2 mm 25 kg	65.6 kg	0.00%
163	109	Omítka weber pas silikon zrnitý 1 mm – 25 kg	41.3 m^2	0.17%
164	109	Omítka sádrová Baumit Glatt L hlazená 1 mm 30 kg	141.4 m^2	0.18%
165	110	Omítka štuková Keraštuk K vnitřní 15 kg	323.5 m^2	0.16%
166	110	Omítka štuková Baumit KlimaPerla vápenná 0,6 mm 25 kg	479.2 m^2	0.31%
167	111	Omítka vápenocementová Baumit UniWhite jemná 0,6 mm 25 kg	1548.8 kg	0.12%
168	111	Omítka vápenocementová Baumit Primo L jádrová 1 mm 40 kg	191.9 m^2	0.13%
169	111	Omítka vápenocementová Baumit UniWhite jemná 0,6 mm 25 kg	202.1 m^2	0.24%
170	113	Beton výplňový Knauf SB 20 30 kg	6.3 m^3	0.24%
171	113	Beton konstrukční Baumit Beton B 20 25 kg	9.1 m^3	0.29%
172	113	Beton C20/25 Cemix 1125 25 kg	11.6 m^3	0.53%
173	113	Beton rychletuhnoucí Knauf BN 30 30 kg	67.1 m^2	0.07%
174	113	Beton drenážní Baumit DrainBeton 4 mm 40 kg	14.5 m^3	0.45%
175	113	Beton C30/37 Sakret BE 04/C30/37 30 kg	6.4 m^3	0.36%
176	113	Beton lehký Baumit PorBeton 20 kg	0.6 m^3	0.16%
177	113	Beton rychletuhnoucí Knauf BN 30 30 kg	2.7 m^3	0.48%
178	113	Směs zálivková Baumit FillBeton 25 kg	9.2 m^3	1.44%
179	114	Cement pro zdění Hranice UNIMALT MC 12,5 25 kg	396.6 kg	0.02%
180	114	Cement portlandský Heidelberg Materials CEM I 42,5 R 25 kg	719.6 kg	0.04%
181	115	Tenkovrstvá malta Ceresit ZU 25 kg	546 kg	0.05%

CB ID	CMPI ID	Item name	Volume	Weight
182	115	Malta vápenocementová Baumit DuoDur 25 kg	0.7 m^3	0.05%
183	115	Malta na pórobeton Baumit PlanoFix 25 kg	1.2 m^3	0.15%
184	115	Malta zdicí a spárovací Baumit Klinker 25 kg	1 m^3	0.14%
185	115	Malta zdicí na pórobeton Baumit PlanoFix 25 kg	0.1 m^3	0.01%
186	115	Malta zdicí Knauf UNI 30 kg	10.6 m^3	0.73%
187	115	Zdicí malta pro vápenopískové a betonové bloky - ZM 920 (40kg/balení)	0.2 m^3	0.04%
188	115	Malta pro lícové zdivo Quick-mix DEK-VM 01-T šedá, 30 kg	0.1 m^3	0.01%
189	115	Malta zdicí KM BETA Sendwix ZM 920 40 kg	0.2 m^3	0.03%
190	116	Plnivo kamenné TopStone Santorini, 2–8	4 m^3	0.62%
191	116	Kamenivo keramické Liapor 8-16 mm 1000 l	8.4 m^3	0.30%
192	116	Kamenivo keramické Liapor 4-8 mm 1000 l	8.4 m^3	0.30%
193	116	Stavební kamenivo frakce 16-22 mm (kačírek)	3413.3 kg	0.09%
194	116	Plnivo kamenné TopStone Bardiglio, 4–7	0.1 m^3	0.08%
195	116	Plnivo kamenné TopStone Giulia šedo-hnědý, 4–8	13.6 m^3	0.33%
196	116	Kamenivo keramické Liapor 1-4 mm 1000 l	3.5 m^3	0.14%
197	116	Písek zásypový DEK jemný 25 kg	6427.7 kg	0.23%
198	116	Písek zásypový DEK hrubý 25 kg	26256 kg	0.91%
199	117	Sádra stavební DEK šedá, 5 kg	10.6 kg	0.00%
200	117	Sádra modelářská Roko bílá, 5 kg	14 kg	0.00%
201	118	Hmota spárovací Cemix 079 FLEX šedá 5 kg	4.6 kg	0.00%
202	118	Hmota spárovací Baumit Baumacol PremiumFuge white 2 kg	2.8 kg	0.00%
203	118	Hmota spárovací Mapei Ultracolor Plus 113 cementově šedá 5 kg	7 kg	0.01%
204	119	Lepidlo cementové Knauf Flexkleber 25 kg	87.8 kg	0.02%
205	119	Lepidlo cementové Mapei Adesilex P9 šedé 25 kg	46.4 m^2	0.04%
206	120	Hydrát vápenný CARMEUSE CL90-S 20 kg	27.1 kg	0.00%
207	120	Vápno bílé hašené Čertovy schody 30 kg	35 kg	0.00%
208	121	Můstek spojovací a ochrana výztuže Sika MonoTop-2001 15 kg	18.2 kg	0.01%
209	121	Hmota stěrková PRIMALEX Beton EFEKT S 4000-N 10 l	140 m^2	0.34%
210	121	Můstek kontaktní Cemix KONTAKT bílý, 24 kg 24 kg	45.7 kg	0.06%
211	122	Stěrka tenkovrstvá Baumit KlimaFino vápenná 20 kg	560.5 m^2	0.20%
212	122	Hmota opravná weberbat 20 kg	143.2 m^2	0.07%
213	123	Postřik sanační cementový webersan podhoz 25 kg	2.6 m^3	0.45%
214	125	Hmota samonivelační weberfloor 4160 25 kg	2.3 m^3	1.70%
215	125	Hmota samonivelační Baumit Nivello 10 25 kg	110.4 m^2	0.37%
216	128	Hrdlo spojovací FF-drän DN 100	5 pcs	0.00%
217	128	T-kus FF-drän DN 100	9 pcs	0.01%
218	129	Trubka drenážní tyčová Opti-drän DN 150 délka 2,5 m	120 m	0.25%
219	129	Trubka drenážní ohebná FF-drän DN 80 délka 50 m	80 m	0.05%
220	130	Lapač střešních splavenin DEK DN 100/125 - šedá	1 pcs	0.00%
221	132	Vpust dvorní s plastovým roštem ACO Self XtraPoint 250×250 mm	2 pcs	0.07%
222	132	Žlab s plastovým roštem ACO Euroline 1,0 m	6 pcs	0.05%
223	134	Silniční obrubník FEROBET rovný 1000/150/250 přírodní	29.2 m	0.06%
224	134	Palisáda betonová DITON POLO přírodní 500×115×115 mm	24.1 m	0.05%
225	135	Betonová dlažba DITON plošná, přírodní 50×500×500 mm	62.2 m^2	0.29%
226	135	Betonová zámková dlažba FEROBET IČKO přírodní, výška 80 mm	72 m^2	0.23%
227	138	Tvárnice svahová DITON DELTA velká přírodní 380×480×250 mm	40 pcs	0.07%
228	140	Plotová brána 3D dvoukřídlá z pozinkované oceli, výška 1530 mm, šířka 4000 mm	1 pcs	0.25%

CB ID	CMPI ID	Item name	Volume	Weight
229	140	Plotová brána dvoukřídlá Pilecký IDEAL II. 3037x1550 mm (Zn+RAL) 6005	2 pcs	0.21%
230	140	Branka jednokřídlá Solid Zn + PVC zelená šířka 1,073 m výška 1,45 m	1 pcs	0.05%
231	141	Pletivo čtyřhranné poplastované Ideal Zn + PVC KOMPAKT antracit výška 1800 m	60 m	0.10%
232	141	Pletivo uzlové pozinkované LIGHT výška 1 800 mm	40 m	0.02%
233	142	Plotovka dřevoplastová DŘEVOplus světlý dub oblouk 15×70 mm	300 pcs	0.52%
234	142	Dřevoplastová plotovka FOREST PLUS, odstín palisander 120x11×3 600 mm	200 pcs	2.14%
235	143	Sloupek pozink DŘEVOplus 60×60 mm 1 m řez	20 pcs	0.10%
236	143	Vzpěra kulatá Ideal Zn průměr 38 mm délka 2,0 m	25 pcs	0.08%
237	143	Deska podhrabová PILHRAB PVC 2450×200×50 mm	10 pcs	0.10%
238	144	Sloupek poplastovaný DŘEVOplus barva šedá 60×60 mm 1,6 m řez	30 pcs	0.24%
239	144	Nosník poplastovaný DŘEVOplus barva hnědá 50×30 mm 3,5 m řez	20 pcs	0.29%
240	145	Tvárnice plotová štípaná DITON A bez fazety cihlová 195×390×190 mm	829 pcs	0.95%
241	145	Stříška plotová DITON universal 30 písková	60 pcs	0.15%
242	153	Cihla plná P40 290×140×65 mm	755 pcs	0.12%
243	153	Cihla plná P20 290×140×65 mm	2500 pcs	0.44%
244	153	Cihla broušená Porotherm 38 Profi Dryfix P10/P8 380×248×249 mm	22 pcs	0.03%
245	153	Cihla akustická HELUZ AKU 20 P15 375×200×238 mm	275 pcs	0.47%
246	153	Akustická cihla HELUZ AKU Z 17,5 P20 175×375×249 mm	343 pcs	0.63%
247	153	Cihla broušená HELUZ FAMILY 50-N doplňková 247×500×166 mm	46 pcs	0.10%
248	153	Cihla broušená Porotherm 24 Profi P10 240×372×249 mm	683 pcs	0.80%
249	153	Cihla POROTHERM 30 P10 247×300×238 mm	1323 pcs	1.18%
250	153	Věncovka broušená HELUZ 8/25 375×80×249 mm	52 pcs	0.05%
251	157	Překlad Porotherm 11,5/7,1/100	1 pcs	0.00%
252	157	Překlad Porotherm 11,5/7,1/200	1 pcs	0.01%
253	157	Překlad Porotherm 7/23,8/100	1 pcs	0.01%
254	157	Překlad Porotherm 7/23,8/125	8 pcs	0.06%
255	157	Překlad Porotherm 7/23,8/150	2 pcs	0.02%
256	157	Překlad Porotherm 7/23,8/175	3 pcs	0.03%
257	157	Překlad Porotherm 7/23,8/250	2 pcs	0.04%
258	157	Překlad nosný Heluz Family 3in1 125	8 pcs	0.90%
259	157	Překlad nosný Heluz 23,8/100	13 pcs	0.07%
260	158	HELUZ panel 1750x600x230 mm	20 pcs	0.95%
261	158	HELUZ panel 3000x900x230 mm	15 pcs	1.83%
262	158	HELUZ panel 3750x900x230 mm	8 pcs	1.22%
263	159	Nosník stropní Porotherm POT 400/902 160×175×4000 mm	4 pcs	0.10%
264	159	Stropní nosník HELUZ MIAKO 5000×160×175 mm	3 pcs	0.10%
265	161	Příčkovka PORFIX P2-500 500×250×50 mm	13 pcs	0.00%
266	161	Příčkovka PORFIX P2-500 500×250×100 mm	5 m^3	0.25%
267	161	Příčkovka Porfix P2-500 75×500×250 mm	7.6 m^3	0.39%
268	161	Tvárnice Porfix Premium P2–400 PDK 500×250×300 mm	33.2 m^3	1.66%
269	162	Překlad nosný KM BETA Sendwix 2DF 250 115×240×2500 mm	6 pcs	0.10%
270	163	Nosník stropní YTONG 1,2/A délka 1 200 mm	3 pcs	0.02%
271	163	Nosník stropní YTONG 1,4/C délka 1 400 mm	40 pcs	0.34%
272	164	Ztracené bednění PORFIX U-profil 500×250×250 mm	286 pcs	0.65%
273	164	U-profil Porfix 200×500×250 mm	680 pcs	0.85%
274	165	Panel balkonový Heluz 1200×230×6400 mm	1 pcs	0.36%
275	204	Vázací drát 0,8 mm nerezový	120 m	0.01%

CB ID	CMPI ID	Item name	Volume	Weight
276	204	Drát vázací ENPRO FeZn 1 mm	24 m	0.00%
277	204	Drát vázací MAX TW1061T	70 m	0.00%
278	204	Drát vázací MAX TW898-EG	60 m	0.00%
279	204	Drát vázací poplastovaný zelený průměr 1,4 mm délka 50 m	150 m	0.00%
280	204	Drát napínací poplastovaný průměr 3,4 mm, délka 52 m	208 m	0.01%
281	204	Napínací drát průměr 3,4 mm délka 26 m s PVC vrstvou	40 m	0.00%
282	204	Drát vázací ENPRO 1,4 mm	50 m	0.00%
283	204	Drát vázací Makita	100 m	0.00%
284	204	Drát vázací MAX TW898-THI	200 m	0.01%
285	204	Vázací drát 0,8 mm poplastovaný	60 m	0.00%
286	208	Vložka cylindrická FAB 1000U4BDNs 29×35 mm	10 pcs	0.21%
287	208	Vložka cylindrická FAB 200RSDNm 40×55 mm	6 pcs	0.08%
288	208	Vložka cylindrická FAB 100RSD/29+35	2 pcs	0.01%
289	210	Kování štítové bezpečnostní FAB BK305/90 madlo/klika F1	5 pcs	0.08%
290	210	Kování rozetové MP Cynthia HR PZ NP/OC hliník	6 pcs	0.04%
291	211	Zámek zadlabací FAB 190/140 /20 L klíč	5 pcs	0.02%
292	211	Zámek zadlabací FAB 190/140/20 P vložkový	6 pcs	0.02%
293	212	Plech krycí DN 80	1 pcs	0.01%
294	213	Čelo žlabu Cu Zambelli 330 mm	4 pcs	0.00%
295	213	Hák žlabový FeZn Zambelli 330/550	160 pcs	0.14%
296	213	Koleno výtokové FeZn Zambelli 72° průměr 80 mm	4 pcs	0.01%
297	213	Kotlík oválný Cu Zambelli 330/100	4 pcs	0.02%
298	213	Klapka výklopná TiZn Zambelli průměr 80 mm	4 pcs	0.02%
299	213	Roh žlabu vnější TiZn Zambelli 330 mm	3 pcs	0.01%
300	213	Hák žlabový obalovaný TiZn Zambelli 330/550	16 pcs	0.02%
301	213	Svod TiZn Zambelli 80 mm délka 1 m	44.4 m	0.13%
302	213	Svod Cu Zambelli 120 mm délka 4 m	4 pcs	0.23%
303	213	Žlab TiZn Zambelli 330 mm délka 6 m	3 pcs	0.07%
304	213	Žlab Cu Zambelli 330 mm délka 4 m	27 m	0.33%
305	214	Ohýbaný lakovaný parapet RS200S-30 PU50 MR539 0,50mm 2bm	10.4 m	0.04%
306	214	Parapet RS250S-30 PU 50 MAX MM133 cihlová	185.2 m	0.79%
307	214	Ohýbaný lakovaný parapet RS200S-30 SP25 RAL3000 0,50mm 2bm	18 m	0.05%
308	215	Plech FeZn tabule tloušťka 0,55 mm	66.8 m	0.59%
309	215	Okapnice DHV RS138S-120 SP25 RAL3005 0,50mm 2bm	30.9 m	0.04%
310	215	Okapový plech RS250S-135 SP25 RAL9007 0,50mm 2bm	15.4 m	0.10%
311	215	Lemování komínu RS416V-60 SP 25 2M295 hnědá	72.2 m	0.25%
312	215	Závětrná lišta RS312SS SP25 RAL5010 0,50mm 2bm	80 m	0.26%
313	219	Tyč betonářská ocelová průměr 18 mm délka 6 m	12 m	0.01%
314	219	Tyč betonářská ocelová průměr 6 mm délka 6 m	24 m	0.00%
315	219	Tyč betonářská ocelová průměr 14 mm délka 6 m	96 m	0.06%
316	219	Tyč betonářská ocelová průměr 8 mm délka 6 m	450 m	0.11%
317	220	Kari síť svařovaná KA 16 oko 100×100 mm drát 4 mm	345.5 m^2	0.46%
318	220	Kari síť svařovaná KD 35 oko 100×100 mm drát 5 mm	139.2 m^2	0.30%
319	220	Kari síť svařovaná KY 50 oko 150×150 mm drát 8 mm	84.9 m^2	0.28%
320	225	Profil výztužný UA 100×40×3000 mm	15 m	0.05%
321	225	Jekl čtvercový 120 x 120 x 3 mm	16 m	0.16%
322	225	Profil výztužný UA 100×40×4000 mm	16 m	0.06%

CB ID	CMPI ID	Item name	Volume	Weight
323	225	Profil výztužný UA 50×40×3000 mm	3 m	0.01%
324	225	Profil výztužný UA 50×40×3500 mm	3.5 m	0.01%
325	225	Profil výztužný UA 50×40×4000 mm	6 m	0.02%
326	225	Profil výztužný UA 75×40×3000 mm	10 m	0.03%
327	225	Profil výztužný UA 75×40×3500 mm	13 m	0.04%
328	225	Profil výztužný UA 75×40×4000 mm	12 m	0.04%
329	225	Profil nosný CD 60×27×3000 mm	50 m	0.03%
330	225	Profil nosný CD 60×27×4000 mm	36 m	0.02%
331	225	Profil nosný CW 100×50×2500 mm	8 m	0.01%
332	225	Profil nosný CW 100×50×2750 mm	12 m	0.01%
333	225	Profil nosný CD 60×27×3000 mm	150 m	0.09%
334	225	Profil nosný CW 100×50×3250 mm	64.2 m	0.07%
335	225	Profil výztužný UA 100×40×3000 mm	72 m	0.26%
336	225	Profil výztužný UA 75×40×2750 mm	4 m	0.01%
337	225	Profil výztužný UA 50×40×3000 mm	9 m	0.02%
338	225	Profil nosný CW 50×50×3000 mm	6 m	0.00%
339	225	Profil obvodový UW 75×40×4000 mm	32 m	0.03%
340	225	Profil nosný CW 50×50×2800 mm	36.7 m	0.04%
341	227	Patka kotevní 16×60×60×4 mm	72 pcs	0.22%
342	227	Spojka křížová CD-K2 DK Mont rovná (15 ks/bal.)	600 pcs	0.06%
343	230	Napínák hák-hák DIN 1480 4.6 M6	12 pcs	0.01%
344	230	Třmen obloukový OMEGA 4.6 8 mm	60 pcs	0.03%
345	230	Kramle tesařská 10×300 mm	92 pcs	0.08%
346	230	Konzola stojánková vnitřní - universal	72 pcs	0.69%
347	231	Hmoždinka univerzální TOX Deco 6×41 mm	350 pcs	0.01%
348	231	Hmoždinka univerzální TOX Trika 12×71 mm	87 pcs	0.01%
349	231	Hmoždinka univerzální TOX Tetrafix 5×25 mm 24 ks	50 pcs	0.00%
350	231	Hmoždinka natloukací zápustná 8×100 mm	100 pcs	0.02%
351	232	Hřebík krovák DEKNAIL Fe 5×150 mm 5 kg	5.6 kg	0.01%
352	232	Hřebík DEKNAIL Fe 2,8×63 mm 5 kg	91.8 kg	0.09%
353	232	Hřebík konvexní 4×40 mm 5 kg	0.4 kg	0.00%
354	232	Hřebík DEKNAIL Fe 2,8×70 mm 5 kg	14 kg	0.01%
355	232	Hřebík DEKNAIL Fe 3,1×90 mm 1 kg	61 kg	0.08%
356	232	Hřebík DEKNAIL Fe 3,1×80 mm 5 kg	1.7 kg	0.00%
357	232	Hřebíky ocelové Rapid High Performance 8×50 mm	3500 pcs	0.01%
358	232	Trn stropní TOX Top 6×65 mm	164 pcs	0.05%
359	232	Hřebíky galvanizované IKO 20 mm 5 kg	1.5 kg	0.00%
360	232	Hřebíky galvanizované IKO 25 mm 5 kg	0.2 kg	0.00%
361	232	Hřebíky galvanizované IKO 30 mm 5 kg	1.4 kg	0.00%
362	232	Hřebíky galvanizované IKO 35 mm 5 kg	24 kg	0.04%
363	232	Hřebík krovák DEKNAIL Fe 7,1×220 mm 5 kg	5.5 kg	0.01%
364	232	Hřebíky galvanizované IKO 20 mm 5 kg	26.8 kg	0.06%
365	232	Hřebíky ocelové Rapid High Performance 8×50 mm	1940 pcs	0.01%
366	233	Závěs přímý CD DK Mont 60/1,0 mm (100 ks/bal.)	290 pcs	0.02%
367	235	Matice nerezová DIN 934 M16	473 pcs	0.04%
368	236	Nýty hliníkové Rapid Standard 3,2×8 mm 100 ks	500 pcs	0.01%
369	237	Podložka nerezová DIN 440 M16	629 pcs	0.02%

CB ID	CMPI ID	Item name	Volume	Weight
370	239	Drát s okem DK Mont 500 mm (100 ks/bal.)	290 pcs	0.02%
371	240	Spony 53 0,7×11,3×14 mm	4000 pcs	0.00%
372	241	Šroub kombi M8 T25 160 mm	69 pcs	0.02%
373	241	Šroub do dřeva Kokeš EDS-H 5×20 mm	405 pcs	0.01%
374	241	Nerezový šroub 13mm, 9546	5850 pcs	0.15%
375	242	Vrut do dřeva DIN 571 M10 6×80 mm	734 pcs	0.02%
376	242	Vrut PAN-HEAD PZ2 5×70 mm	438 pcs	0.01%
377	243	Tyč závitová DIN 975 4.8 M6 1 000 mm	60 pcs	0.05%
378	303	Barva fasádní weberton silikát bílá 25 kg	400 m^2	0.24%
379	303	Barva fasádní Baumit SilikatColor probarvená 14 l	315 m^2	0.35%
380	304	Lazura středněvrstvá Remmers UV+ bezbarvý, 0,75 l	116.6 m^2	0.06%
381	304	Olej UV ochranný Osmo 426 extra modřín, 2,5 l	60 m^2	0.05%
382	304	Lazura olejová ochranná Osmo 701 bezbarvá 2,5 l	345.4 m^2	0.46%
383	304	Olej tvrdý voskový Osmo Rapid 3240 bílý 2,5 l	120 m^2	0.26%
384	305	Malba interiérová Primalex Standard bílá, 40 kg	657.8 m^2	0.04%
385	305	Malba interiérová HET Hetline San Active bílá, 15 kg	613.8 m^2	0.16%
386	305	Malba interiérová Caparol Primacryl bílá, 25 kg	300 m^2	0.26%
387	307	Nátěr epoxidový Cemix Cempaint FLOOR EPW šedý, 10 kg	20 m^2	0.15%
388	308	Barva samozákladující Rokosil akryl 3v1 RK 300 červ. rum. 0,6 l	10 m^2	0.03%
389	308	Barva epoxidová Stachema Sinepox S 2321 na vany bílý	7.2 m^2	0.01%
390	308	Barva samozákladující Rokosil akryl 3v1 RK 300 červ. rum. 0,6 l	22 m^2	0.00%
391	308	Email vrchní Rokoemail S2013 červenohnědá, 3 l	5 m^2	0.01%
392	308	Email vrchní Rokoemail S2013 hnědá střední, 0,6 l	4.1 m^2	0.01%
393	311	Páska maskovací Masq béžová 36 mm/40 m	775.4 m	0.01%
394	312	WPC impregnace pro terasová prkna Woodplastic, 2,5 lt	67.1 kg	0.48%
395	313	Lepidlo na dřevo Den Braven Wood Fix D3 500 g	3.9 kg	0.01%
396	313	Lepidlo polyuretanové INSTA-STIK 750 ml	0.8 kg	0.00%
397	314	Pěna montážní Soudal Monton pistolová 750 ml	3 pcs	0.01%
398	314	Pěna montážní Soudal PERI BOND 750 ml	154 m	0.05%
399	315	Penetrace Cemix ST Color 24 kg 24 kg	20 kg	0.03%
400	315	Nátěr penetrační Ceresit CT 19 Supergrip 5 kg	55.1 m^2	0.07%
401	316	Superplastifikátor Sika 1 l	12 kg	0.03%
402	316	Plastifikátor GABOTHERM Gabolith GTF EZ 10 l	26.5 kg	0.04%
403	317	Organické rozpouštědlo k ředění lepidel a čištění	37.4 kg	0.07%
404	317	Ředidlo nitro Severochema C 6000 – 4 l	0.9 kg	0.00%
405	317	Systémové rozpouštědlo ALKORPLUS 81025 tetrahydrofuran k PVC-P fóliím 1 l	1.2 kg	0.01%
406	318	Tmel silikonový HACO USB 60 G	95 m	0.01%
407	318	Akrylový tmel Soudal Acryrub 600 ml, bílý	3.1 kg	0.01%
408	318	Konstrukční tmel 25D (280 ml/bal) šedý	4.8 kg	0.06%
409	318	Tmel spárovací Rigips Max 5 kg	40 kg	0.02%
410	318	Silikon neutrální sanitární +S Soudal transparentní 280 ml	12 pcs	0.03%
411	318	Tmel spárovací DEKFINISH 5 kg	98.5 kg	0.02%
412	405	Dřez Jika DORIS 59 cm s přepadem	1 pcs	0.05%
413	405	Dřez nerezový Sanela SLUN 04E, 24 V DC	2 pcs	0.58%
414	405	Dřez nerezový Sanela SLUN 15	1 pcs	0.22%
415	410	Závěsný bidet Jika PURE včetně instalační sady Easy fit	1 pcs	0.05%
416	413	WC závěsné Jika Deep by Jika 70 cm ZTP	1 pcs	0.06%

CB ID	CMPI ID	Item name	Volume	Weight
417	413	WC kombinované Jika Lyra Plus vodorovný odpad	2 pcs	0.09%
418	413	Závěsný klozet Jika DEEP BY JIKA 70 cm, pro ZTP	2 pcs	0.12%
419	414	Dvojumyvadlo Jika MIO-N 130 cm s otvorem pro baterii	1 pcs	0.13%
420	414	Umyvadlo Laufen PRO S 55 c, s otvorem	1 pcs	0.04%
421	414	Umyvadlo Ravak Chrome 600 mm s otvorem	2 pcs	0.13%
422	414	Umývátko Jika DEEP BY JIKA 50x23 cm, otvor pro baterii vlevo	2 pcs	0.02%
423	417	Vanička sprchová Kaldewei CONOFLAT 781-1 800×1000×32 mm smaltovaná ocel	1 pcs	0.28%
424	419	Stěna pevná Ravak CPS 900 mm bright alu/transparent	2 pcs	0.20%
425	419	Dveře posuvné třídílné SanSwiss TOPS3 900 mm, Aluchrom, čiré	1 pcs	0.13%
426	419	Stěna boční SanSwiss TOPF 900 mm, Aluchrom, čiré	1 pcs	0.09%
427	421	Vana akrylátová Jika Cubito 180×80 cm	1 pcs	0.10%
428	421	Vana akrylátová Roth Kubic Neo 160×70 cm	1 pcs	0.10%
429	422	Nohy vanové Kaldewei 5030	1 pcs	0.02%
430	422	Nohy k vanám EXCLUSIVE	1 pcs	0.01%
431	427	Vtoková armatura s těsněním Sanela SLA 02	5 pcs	0.01%
432	429	Kohout radiátorový Slovarm VE-4523A DN 20	3 pcs	0.04%
433	429	Termostatický ventil IMI samotížný DN 20 přímý	2 pcs	0.02%
434	429	Ventil termostatický LUXOR 3/4" přímý	3 pcs	0.02%
435	431	Šroubení mosazné 3/4", pár	6 pcs	0.02%
436	431	Uzavírací šroubení IMI Regutec DN 10 rohové	1 pcs	0.00%
437	432	Kohout kulový plyn 3/4" MF s pákou	1 pcs	0.00%
438	432	Kohout kulový voda 3/8" FF s motýlkem	3 pcs	0.00%
439	432	Redukční ventil D03-3/4" C	1 pcs	0.01%
440	432	Ventil přístrojový Novaservis CF3017 1/2"×3/4"	5 pcs	0.01%
441	435	Hadice plynová Sievert 7173-41 4 m	12 m	0.07%
442	438	Čerpadlo cirkulační Wilo Star-Z NOVA T, 230 V	1 pcs	0.06%
443	440	Nádrž na dešťovou vodu ECO 5000 l	1 pcs	0.38%
444	441	Mezikus spojovací pro 2 rychlospojky ENPRO	124 pcs	0.03%
445	441	Rychlospojka Rosa 3/4″ mosaz	80 pcs	0.14%
446	444	Plastová odpadní trubka HTEM DN 50, délka 1000 mm	20.2 m	0.02%
447	444	Plastová odpadní trubka HTEM DN 100, délka 1000 mm	21.1 m	0.06%
448	444	KGEM trubka s hrdlem pro kanalizaci DN 100, délka 2000 mm	28.2 m	0.06%
449	444	KGEM trubka s hrdlem pro kanalizaci DN 150, délka 2000 mm	23.2 m	0.09%
450	445	Trubka měděná 18×1 mm × 2,5 m, polotvrdá	9 m	0.02%
451	445	Trubka z uhlíkové oceli 18 mm, 3 m	12 m	0.01%
452	445	Trubka měděná 15×1 mm × 5 m, polotvrdá	58 m	0.11%
453	445	Trubka měděná 15×1 mm × 5 m, polotvrdá	58 m	0.11%
454	445	Trubka měděná 42×1,2 mm × 5 m, tvrdá	9 m	0.07%
455	446	Plastové potrubí LDPE PE040 20x3,0 PN10 x 150 bm	62 m	0.02%
456	446	Trubka PP-RCT CARBO D25×3,5 mm × 4 m	100 m	0.13%
457	446	Trubka PP-RCT UNIBETA D32×3,6 × 4 m	9 m	0.01%
458	447	Víčko lisovací Sanha Cu Press 6301 15 mm	20 pcs	0.04%
459	447	Rozeta nízká Alca A98 DN 100	8 pcs	0.00%
460	447	Kus prodlužovací Alcadrain A4000 DN 32	2 pcs	0.00%
461	447	Přímá manžeta Alcaplast A99	4 pcs	0.00%
462	448	Hlavice termostatická DEK TRV1	1 pcs	0.00%
463	448	Termostatická hlavice IMI DX	2 pcs	0.01%

CB ID	CMPI ID	Item name	Volume	Weight
464	449	Objímka dvoušroubová 20-24 mm (1/2") M8	35 pcs	0.01%
465	449	Třmen s navrtávkou Gebo ANB 2 1/2"×1"	9 pcs	0.12%
466	450	Žlab podlahový ke stěně HACO PLZ 750 SM	3 pcs	0.20%
467	450	Odtok vodorovný Wedi Fundo DN50	2 pcs	0.03%
468	450	Žlab podlahový Alcaplast APZ8-750M Simple včetně roštu	4 pcs	0.13%
469	450	Žlab sprchový Alcaplast APZ4-750 Flexible	2 pcs	0.12%
470	450	Žlab sprchový nerezový Alca APZ13-750 Modular	1 pcs	0.04%
471	450	Výpust umyvadlová Alca A31 5/4" s nerezovou mřížkou	4 pcs	0.00%
472	453	Redukce KGR DN 125/100	2 pcs	0.00%
473	453	Přesuvka HTU DN 50 pro dodatečné spojování trubek	13 pcs	0.01%
474	453	Nátrubek lisovací Sanha Cu Press 6270 15 mm	16 pcs	0.01%
475	453	Lisovací přechodka Sanha Cu SOLAR 18×3/4" vnější 13243G	18 pcs	0.03%
476	453	Spojka Gebo PPS 50×6/4", vnější závit	11 pcs	0.02%
477	453	Redukce krátká HTR DN 40/32	12 pcs	0.01%
478	453	Nátrubek mosazný 3/4"	9 pcs	0.01%
479	454	Sifon pračkový podomítkový Alcaplast APS3P nerez s přivzdušněním	3 pcs	0.02%
480	454	Dřezový sifon Alcaplast A441P DN50/40 s nerezovou mřížkou DN70 a přípojkou	3 pcs	0.01%
481	454	Sifon umyvadlový celokovový Alcaplast A432 DN 32 tvar "U", s převlečnou maticí	2 pcs	0.01%
482	454	Sifon umyvadlový Alcaplast A434 DN 40 prostorově úsporný s převlečno	1 pcs	0.00%
483	454	Sifon vanový s ovládáním Alcaplast A55K kov	3 pcs	0.03%
484	455	Set těsnění KGGA DN 125	80 pcs	0.04%
485	455	Těsnění hadicových spojek 24×16×2 mm, 7ks/bal., velikost 1/2"	280 pcs	0.02%
486	455	Těsnění mezi hadici a spojku Regulus	15 pcs	0.01%
487	456	Vodoměr mokroběžný vícevtokový IBRF SV DN25	3 pcs	0.12%
488	458	Kus čisticí KGRE DN 100	3 pcs	0.02%
489	458	HTB koleno s hrdlem pro odpadní potrubí, DN 50, úhel 45°	10 pcs	0.00%
490	458	Koleno HTB DN 32 - 87°	10 pcs	0.00%
491	458	Koleno jednoduché PPs DN 60 45°	10 pcs	0.03%
492	459	Koleno 90° lisovací Sanha Cu Press Gas 11090G 28×1" vnitřní závit	3 pcs	0.02%
493	459	Koleno mosazné 90° 1/2" MF	7 pcs	0.01%
494	459	T-kus mosazný 1/2" FFF	24 pcs	0.03%
495	459	T-kus redukovaný pozinkovaný 3/4"×1/2"	19 pcs	0.01%
496	459	Oblouk 90° lisovací Sanha Cu Press Gas 10002A 15 mm FF	30 pcs	0.04%
497	460	Koleno odpadu komplet Alcaplast M906 90/110 mm	5 pcs	0.01%
498	460	T-kus Gebo PPS 20×20×20 mm	20 pcs	0.02%
499	460	Křížení PP-R D25	18 pcs	0.01%
500	461	Beztlaková stojánková kohoutková baterie Clage SNT	1 pcs	0.02%
501	462	Baterie bidetová Kludi Pure&Easy 375330565 chrom	2 pcs	0.07%
502	463	Baterie dřezová stojánková Novaservis EDGE 36713,GRS granit–písek	1 pcs	0.05%
503	463	Baterie dřezová stojánková RAV-Slezák Sázava SA006.0/13S chrom-šedá	1 pcs	0.02%
504	463	Baterie dřezová stojánková Novaservis METALIA ECO 57014,0E, chrom	1 pcs	0.03%
505	464	Baterie podomítková RAV-Slezák Colorado CO186K chrom s přepínačem	1 pcs	0.04%
506	464	Baterie sprchová termostatická Grohe GROHTHERM 800 COSMOPOLITAN	3 pcs	0.17%
507	465	Baterie umyvadlová stojánková Grohe EUROSMART 23323001 M s řetízkem	2 pcs	0.05%
508	465	Baterie umyvadlová stojánková Jika Cube H3111W10045101 chrom	3 pcs	0.07%
509	465	Baterie umyvadlová stojánková RAV-Slezák Victoria VI226.0 chrom	1 pcs	0.01%
510	466	Baterie vanová nástěnná Grohe EUROSTYLE COSMOPOLITAN 150 mm chrom	1 pcs	0.04%

CB ID	CMPI ID	Item name	Volume	Weight
511	466	Baterie vanová podomítková Grohe EUROSMART 33305002	1 pcs	0.03%
512	468	Nádoba expanzní Reflex Refix DE 18/10 neprůtočná	1 pcs	0.02%
513	468	Nádoba expanzní Reflex S 50/10	1 pcs	0.04%
514	470	Elektrokotel Bosch Tronic 5000 H 45 kW	1 pcs	0.46%
515	471	Kamna krbová Haas+Sohn EBOLI	1 pcs	0.51%
516	472	Kotel na tuhá paliva Enbra TP-EKO 25 MINI	1 pcs	0.77%
517	473	Elektrický průtokový ohřívač Clage MCX 7 6,5 kW 400 V	1 pcs	0.09%
518	474	Plynový kondenzační kotel WOLF FGB-K-24 kombi	1 pcs	0.62%
519	475	Komínová sestava Schiedel UNI Advanced DN 160 výška 7 m	1 pcs	0.38%
520	475	Přechod stěnový DN 80/125-DN 80	3 pcs	0.14%
521	475	Komínek ležatý jednoduchý DN 80	3 pcs	0.05%
522	475	Komínová sestava Schiedel Absolut DN 160 výška 6 m	1 pcs	0.51%
523	476	Trubka odkouření černá DN 200/1000 mm	2 pcs	0.03%
524	476	Zděř odkouření černá DN130	2 pcs	0.00%
525	479	Rohož topná Fenix LDTS 160 13,3 m2	2 pcs	0.27%
526	479	Sada topné folie Fenix EcoFilm set 2 m 80 W/m2	1 pcs	0.02%
527	480	Okruh topný Fenix MADPSP 40 W/m 47 m 1 887 W	2 pcs	0.07%
528	480	Kabel topný samoregulační V-systém SR PRO 20 20–37 W/m (metráž)	1 pcs	0.00%
529	481	Konvektor podlahový Boki InFloor F1S 200×1000×75 mm s ventilátorem 24 V DC	1 pcs	0.11%
530	484	Radiátor deskový Stelrad COMPACT ALL IN 22 (500×1000 mm)	3 pcs	0.12%
531	484	Radiátor deskový Stelrad NOVELLO 21 (500×1000 mm)	4 pcs	0.18%
532	484	Radiátor deskový Stelrad VERTEX PLAN 11 (H2000×400 mm)	7 pcs	1.03%
533	485	Radiátor trubkový KDO 750×960 mm bílá	2 pcs	0.05%
534	485	Radiátor trubkový elektrický KH-E 750×1330 mm bílá	1 pcs	0.08%
535	485	Radiátor trubkový KM MARABU 750×1815 mm bílá	1 pcs	0.08%
536	490	Kryt k ventilátoru ICON 30, antracit	3 pcs	0.04%
537	490	Mřížka větrací se sítí a krytkou HACO VM K 200×300 mm bílá	5 pcs	0.01%
538	490	Mřížka větrací horizontální Regulus 60×200 mm	1 pcs	0.00%
539	490	Ventil talířový Regulus DN 125	2 pcs	0.00%
540	492	Tvarovka T HACO CTP 3× 110×55 mm	2 pcs	0.00%
541	492	Trubka čtyřhranná plastová Regulus 60×200 mm délka 1,5 m	5 m	0.02%
542	492	Odbočka jednoduchá Regulus DN 125/125 90°	3 pcs	0.01%
543	492	Hadice tepelně izolovaná Regulus 127 mm délka 10 m	2 pcs	0.04%
544	494	Ventilátor domovní Klimatom Primo base 100 AH, 12 W, 230 V	1 pcs	0.01%
545	494	Ventilátor domovní Blauberg Auto 125 T, 22 W, 230 V, IP 24	2 pcs	0.03%
546	495	Zásobník teplé vody Thermona Therm 60/S	1 pcs	0.23%
547	503	Svítidlo LED s dálkovým ovládáním Emos ZM5166 36 W	2 pcs	0.06%
548	503	Svítidlo LED Ecolite Lada 2 18 W 1 550 lm	6 pcs	0.05%
549	503	Svítidlo LED s dálkovým ovládáním Emos Exclusive 18 W CRI>95	2 pcs	0.05%
550	503	Svítidlo LED Greenlux Fenix Square 12 W 3 800 K	3 pcs	0.02%
551	503	Svítidlo LED Modus BRSB 27 W	6 pcs	0.08%
552	503	Svítidlo LED Ecolite Victor 18 W 4 100 K	6 pcs	0.13%
553	506	Svítidlo E27 Eglo Bolovone 60 W	2 pcs	0.03%
554	506	Svítidlo LED Ledvance 14 W 4 000 K	3 pcs	0.02%
555	508	Svítidlo LED Philips Ledinaire 30 W	4 pcs	0.06%
556	508	Svítidlo LED Pila WT007C 54 W 5 400 lm	2 pcs	0.02%
557	508	Svítidlo LED Trevos Prima 37 W 5 500 lm	2 pcs	0.04%

CB ID	CMPI ID	Item name	Volume	Weight
558	509	Svítidlo E27 Eglo Rocamar 40 W	4 pcs	0.06%
559	509	Svítidlo LED UGR Modus LLLX6000 57 W 6 200 lm	3 pcs	0.07%
560	510	Transformátor pro LED osvětlení 16W, Skoff	3 pcs	0.02%
561	510	Stmívač LED otočný 3–60 W Emos	2 pcs	0.01%
562	514	Pásek LED Led-Pol 12 V 14,6 W/m růžová	3 pcs	0.02%
563	514	Pásek LED Panlux 24 V 6 W/m 4 000 K	2 pcs	0.01%
564	519	Svítidlo LED Philips CoreLine Malaga, 83 W, 9000 lm, 4000 K	2 pcs	0.12%
565	520	Svítidlo E14 sloupek se zásuvkou Panlux Gard 36 60 W	4 pcs	0.04%
566	520	Svítidlo LED Ecolite Nela 24 W	6 pcs	0.06%
567	523	Drát zemnicí FeZn pr. 10 mm (50kg/bal = 80,5bm)	120 kg	0.16%
568	523	Pás zemnicí FeZn 30×4 mm (25kg/Bal)	30 kg	0.04%
569	524	Podpěra vedení Kovoblesk PV15 Beta, nerez N-V2A	25 pcs	0.02%
570	524	Svorka zemnicí , SR 03 S	6 pcs	0.00%
571	524	Držák jímače a ochranné trubky, DJDe 60	4 pcs	0.00%
572	525	Tyč zemnicí, ZTP 2	10 pcs	0.09%
573	526	Krabice do sádrokartonu, KPR 68/71L	10 pcs	0.66%
574	526	Krabice pod omítku s víčkem a svorkovnicí, KR 97/5 KA	22 pcs	0.05%
575	526	Krabice instalační Schneider Unica, 2× 6 moduly	12 pcs	0.03%
576	528	Jistič Schneider A9F03116 6 kA B 16 A	15 pcs	0.02%
577	528	Svodič přepětí T2 Eaton SPCT2-280/4	2 pcs	0.09%
578	528	Jistič OEZ LTS-16B-1 10 kA B 16 A	10 pcs	0.02%
579	529	Vložka nožová pojistková OEZ PNA000, 500 V, gG, 40 A	5 pcs	0.00%
580	530	Nástavný rámeček, NR 68/6 ZB	6 pcs	0.01%
581	531	Rozvodnice Hager Golf VS218PD, nástěnná, IP 30, plast	3 pcs	0.05%
582	531	Rozvodnice prázdná plastová SEZ CZ P-BOX, 300×400×220 mm	7 pcs	0.11%
583	533	Stykač Schneider Acti 9 iCT A9C20732, 25 A, 230 V, 2 S	3 pcs	0.03%
584	533	Vypínač hlavní Eaton IS-32/3, 3pól, 32 A, 240/415 V	4 pcs	0.02%
585	535	Kabel koaxiální EMOS, úhlová vidlice 10 m	40 m	0.01%
586	535	Kabel sdělovací, TCEPKPFLE 5×4×0,8	60 m	0.02%
587	536	Spojka lisovací Cimco 4 mm2 19 mm 100 ks/bal.	25 pcs	0.00%
588	536	Svěrka koncová RSA L 15	36 pcs	0.01%
589	537	Kabel pryžový H07RN-F 5G2,5 (metráž)	182 m	0.19%
590	537	Kabel CYKY-J 3× 1,5 RE metráž	400 m	0.10%
591	537	Kabel CYKYLo -J 3× 1,5 RE 100 m/bal.	420 m	0.11%
592	538	Příchytka kabelová s hřebíkem pr. 5 mm	70 pcs	0.07%
593	539	Vodič CY H07V-U černá 6 mm2	60 m	0.02%
594	542	Datová zásuvka Keystone RJ 45 CAT6	8 pcs	0.02%
595	544	Ovládač zapínací dvojitý řazení 1/0+1/0 Decento černá	19 pcs	0.31%
596	544	Přepínač střídavý řazení 6 Element bílá	14 pcs	0.03%
597	544	Spínač dvojpólový Schneider Sedna Design řazení 2 bílá	25 pcs	0.05%
598	546	Dvojzásuvka natočená Tango, ochranné kolíky, clonky, černá	15 pcs	0.04%
599	546	Zásuvka jednonásobná s víčkem ABB Tango IP 44	45 pcs	0.11%
600	546	Zásuvka jednonásobná s clonkami Levit bílá	17 pcs	0.03%

Josef Štefl

High Frequency Price Index of Construction Materials: Nowcasting of producer prices

Master's thesis

2024, Prague
Charles University, Institute of Economic Studies
Typesetting Josef Štefl
Number of pages 101
Number of eligible characters 144,376