Working Paper Series 14/2023

Ace in Hand: The Value of Card Data in the Game of Nowcasting

Tomáš Adam, Jan Bělka, Martin Hlůže, Jakub Matějů, Hana Prause, Jiří Schwarz

Czech National Bank —— Working Paper Series —— 14/2023

CZECH NATIONAL BANK

www.cnb.cz

Reviewed by:	Matías Pacce (Bank of Spain)
	Róbert Ambriško (Czech National Bank)
Project Coordinator:	Michal Franta

Ace in Hand: The Value of Card Data in the Game of Nowcasting

Tomáš Adam, Jan Bělka, Martin Hlůže, Jakub Matějů, Hana Prause, Jiří Schwarz*

Abstract

We use Mastercard card payments data to nowcast turnover in Czech retail sales and services. We show that an index based on this data tracks surprisingly well the official retail sales data released by the Czech Statistical Office (CZSO) more than a month later. We further show that the card payments data not only helps in backcasting Czech retail sales after the end of the month, but also provides valuable information for the nowcast as soon as three weeks into the ongoing month. That is six to seven weeks ahead of the official release. To illustrate the usefulness of our method, we show that we would have been able to backcast, with reasonable accuracy, the sharp drop in retail sales that occurred at the outbreak of the first wave of covid-19 in Czechia in March 2020 four weeks before the March data was released by the CZSO.

Abstrakt

Zkoumáme využití transakčních dat z karet Mastercard pro krátkodobou predikci tržeb v maloobchodě a službách. Ukazujeme, že index sestavený z těchto dat překvapivě přesně odpovídá datům, která vydává Český statistický úřad (ČSÚ) o více než měsíc později. Ukazujeme, že index z karetních dat nejen zlepšuje predikci českých maloobchodních tržeb po konci příslušného měsíce („backcast"), ale poskytuje také hodnotnou informaci o vývoji v daném měsíci již po uplynutí tří týdnů. To je šest až sedm týdnů před zveřejněním oficiálních dat ČSÚ. K ilustraci přínosu tohoto přístupu ukazujeme, že bychom byli schopni na jeho základě s dostatečnou přesností zachytit prudký pokles maloobchodních tržeb v první vlně covid-19 v ČR v březnu 2020, a to čtyři týdny před zveřejněním březnových dat ČSÚ.

JEL Codes: E21, E27.

Keywords: Card payments data, household consumption, household demand, nowcasting, retail sales, sales in services.

* Tomáš Adam, Czech National Bank, tomas.adam@cnb.cz
Jan Bělka, Mastercard, jan.belka@mastercard.com
Martin Hlůže, Mastercard, martin.hluze@mastercard.com
Jakub Matějů, Czech National Bank, jakub.mateju@cnb.cz
Hana Prause, Mastercard, hana.prause@mastercard.com
Jiří Schwarz, Anglo-American University, j.schwarz.sba@aauni.edu (corresponding author)

The authors would like to thank Róbert Ambriško, Matías Pacce, and Michal Franta, as well as seminar participants at the Czech National Bank for comments and suggestions. The views expressed in this paper are those of the authors and not necessarily those of the Czech National Bank.

1. Introduction

Macroeconomic policy makers need accurate and timely information on current developments in the economy. As the covid-19 pandemic has highlighted, relying on the official releases of national statistical offices, which are published with significant delays, may be problematic when the economic environment is rapidly changing. Monetary policy is adjusted eight times a year in Czechia. The six- or seven-week time advantage provided by the nowcast of retail sales and sales in services that we introduce in this analysis may prove crucial in the event of non-cyclical shocks to the economy. With the benefit of hindsight, we can say that the ability to monitor the response of household demand in near real time would have been a major advantage for monetary policy makers in the context not only of the introduction of covid-19 lockdowns, but also of the lifting of those lockdowns in mid-2021. In normal economic times, the added value of such a time advantage is limited.

The field of macroeconomic nowcasting pioneered by Stock and Watson (1989) has made great progress in capturing macroeconomic turning points in time and refining the real-time assessment of current economic developments. Nowcasting applications on Czech data focus mainly on GDP growth and include Arnoštová et al. (2011), Rusnák (2016), Franta et al. (2016), and Adam and Novotný (2018). Adam et al. (2021) use alternative high-frequency data such as electricity consumption, tolls collected on Czech highways, and Google trends to construct a weekly tracker of Czech GDP dynamics. Ambriško (2022) uses daily fiscal spending data to nowcast the wage bill, private consumption, and GDP.

Recently, the increasing availability of big granular datasets in real time has opened up space for a number of potential nowcasting applications. Specifically, card payments data are gaining prominence in nowcasting retail sales and personal consumption expenditure. As the share of card payments in retail transactions grows, card payments data are becoming a timely source of information about underlying aggregate demand trends.

One of the first published applications is Verbaan et al. (2017), who use card payments data to improve the nowcast of household consumption in the Netherlands. Galbraith and Tkacz (2018) illustrate that card payments data can reduce nowcast errors for retail sales and GDP in Canada. Similarly to our paper, García et al. (2021) show that card payments data can closely track the official retail sales data, but without the publication lag. Their dataset is limited to clients of one specific bank (BVVA), yet they still achieve results valid for the entire Spanish economy. Barlas et al. (2021) extend this exercise to nowcast Turkish investment and consumption. Aastveit et al. (2020) use Norwegian payments data in a mixed-frequency MIDAS model to nowcast household spending.

The covid-19 pandemic has further sparked the demand for such applications. Chapman and Desai (2021) show that the gains from using retail payments data in Canada were huge during the covid-19 shock but limited in calm periods. Even though we also focus mainly on the covid-19 period, our analysis goes beyond this finding, suggesting that card payments data adds value not only in turbulent times such as the covid-19 lockdowns, but through the whole economic cycle. Using similar data to ours, Alcedo et al. (2022) explore the universe of global Mastercard data to illustrate how e-commerce spending trends have changed after covid-19 across different sectors.

To the best of our knowledge, our study is the first to address nowcasting of retail sales and sales in services. We focus on the ability of aggregated and anonymized Mastercard card payments data ("MC data") to improve the accuracy of nowcasts compared to benchmark models using only official Czech Statistical Office data ("CZSO data"). We show that an index built from the MC data tracks surprisingly well the CZSO data released five to six weeks after the end of the month. This finding holds not only for the post-2020 period, characterized by repeated covid-19 lockdowns, but also for the previous years of a standard economic cycle. As we are solely interested in the added value of the MC data, we do not look for other high-frequency variables, such as monthly VAT returns, which could also potentially be exploited.

To improve the fit, we carefully match the Mastercard Merchant Category Codes to the corresponding CZSO categories without imposing or estimating any additional weighting schemes. We further show that the MC data not only help in backcasting Czech retail sales after the end of a month when the payment data is complete, but also provides valuable information for nowcasting as soon as three weeks into the ongoing month. This is roughly six to seven weeks ahead of the official CZSO release. We also extend this analysis to draw implications about the value of Mastercard data for nowcasting quarterly GDP, which is limited by the nowcasting value of the official CZSO data itself.

For monetary policy purposes, it is considered best practice to forecast seasonally adjusted values at constant prices to describe the underlying fundamental evolution of the business cycle. However, we focus on the ability to nowcast retail sales following exogenous non-cyclical shocks. In such periods, seasonal patterns may change, which we confirm for the covid-19 pandemic in 2020–2021. As a result, the use of unadjusted series is less problematic or even preferable for this use case compared to the situation of a model developed for nowcasting macroeconomic variables over a standard cycle.

The paper is organized as follows. The next section describes the differences between the CZSO data and the MC data and the construction of a retail sales index based on card payments data. Section 3 describes the methodology for testing the backcasting and nowcasting properties of the MC data. Then we present the results and apply the estimated models to the covid-19 period to show the usefulness of timely information in periods of abnormal macroeconomic volatility. The last section concludes.

2. Data

2.1 Monthly Indicators Published by the Czech Statistical Office

The Czech Statistical Office ("CZSO") publishes indices of monthly turnover in retail and selected market services as a measure of economic activity in the economy.[1] The dominant source of data is the relevant monthly survey conducted by the CZSO. Since 2020, probably in connection with the

---

[1] We use the term "retail sales" to denote a general monthly series capturing the indices of retail sales or turnover in services published by the CZSO. Additionally, the term can refer to either the aggregate time series or a subcomponent thereof.

covid-19 pandemic, the survey data have been accompanied by selected administrative data, such as monthly VAT tax returns.

The sample population consists of economically active units classified into selected CZ-NACE categories: *wholesale and retail trade and repair of motor vehicles and motorcycles* (NACE 45), *retail trade, except of motor vehicles and motorcycles* (47), *transportation and storage* (49–53), *accommodation and food service activities* (55–56), *information and communication* (58–63), *real estate activities* (68), *professional, scientific and technical activities* (69–71 + 73–74), and *administrative and support service activities* (77–82). From 2009 to 2019, units with an annual turnover of at least CZK 250 million as well as units with at least 50 employees were included in the sample. Since 2020, the criteria for sample selection have been set for each CZ-NACE category individually. The probability of a unit being included is given by the contribution of its sales to the total sales in the respective CZ-NACE category.

The standard publication date is the 36th day after the end of the observed month plus up to another seven days depending on the specific month. Sales in market services for the quarter-ending months are published with an additional lag of a couple of days. Therefore, the data is usually available five to six weeks after the end of the observed month. The published data is then revised and corrected together with the publication of the following month's data. The final revision of the whole-year monthly data takes place after the publication of the December data, usually in March of the following year.

The level of disaggregation varies. In some cases, data are provided on the four-digit level of classification, such as *retail sale in non-specialised stores with food, beverages or tobacco predominating* (NACE 47.11). Due to strong seasonality, the data are published in both seasonally and calendar unadjusted and adjusted form to allow for month-on-month comparisons. The adjustment is done using the TRAMO/SEATS method, but no details regarding the specification used are published.

The data for monthly sales in retail and selected market services is available from January 2000, which constitutes the beginning of our training period for the benchmark model.

2.2 Mastercard Card Payments Data[2]

The analysis leverages anonymized Mastercard data. The card payments data consist of financial data on payments made in the economy via electronic payment cards. Since electronic cards are commonly used nowadays to pay for goods and services, they represent a significant share of consumption in the economy.[3]

The payment process involves a transfer of funds for a specific transaction (purchase) between the cardholder's account and the merchant's (business) account. The card payments data contains information about the date and time of the transaction, its amount and currency, the payment

---

[2] For this study, the MC data was used in aggregated and anonymized form and only in the transformations described (i.e., calculated monthly indices).

[3] According to the SBK (http://www.bankovnikarty.cz/pages/czech/profil_karty.html), retail payments account for 64% of the total volume of card transactions, the remaining 36% being cash withdrawals from ATMs.

interaction method (e.g., via contactless technology, via magnetic stripe), the type of payment card, cardholder presence in the transaction process, and details about the issuer's bank, the merchant's business (including location), and the acquirer's bank (the one operating the payment terminal). It does not include any information about individual consumers or cardholders. Mastercard collects data about cleared transactions in a batch system. Insights for analytics and further applications are technically available with a lag of up to seven days after the original transaction date and are not subject to revisions.[4] This offers a significant advantage over traditional retail sales data sources based on manual data collection or surveys.

To estimate the value of retail sales and services in the Czech economy, we included processed and reported transactions (both online and offline) from all Mastercard issued cards worldwide with merchant locations in the Czech Republic. For this analysis, we excluded all cash withdrawals.[5] Each merchant location has its own MCC (Merchant Category Classification) code. The MCC code is used to classify a merchant (business) according to the types of goods or services it provides. There are about 1300 MCC codes available. To assess the relevance of the transaction data for the estimation of total retail sales and services in the Czech economy, we focused on the definitions of the individual MCC codes to ensure that they are consistent with the existing CZSO data composition.

We went through the whole card payments data sample and distributed the CZ-NACE codes – at the lowest level of aggregation possible – to the individual MCC categories. As the MC data cover a wider set of sectors than the CZ-NACE codes (for retail and services), matching and filtering of the data was necessary. The matching process resulted in the exclusion of selected category types which are not included in the CZSO data structure. Then we aggregated the MC data to match the data and structure published by the CZSO as closely as possible and calculated the month-on-month changes for the selected NACE categories. No additional information was used or available for the analysis.

Figures C1 and C2 in Appendix C visualize the shares of the individual subcategories of *retail trade, except of motor vehicles and motorcycles* (NACE 47) and all services (NACE 49–82) within the CZSO and MC data. The difference in the shares of the CZ-NACE subcategories mainly for the NACE 47 category is negligible, as card penetration in retail is already high. The differences are higher in the case of services, where restaurants and accommodation (NACE 55–56) account for a large proportion of transactions (almost 55%), whereas their weight in the CZSO data is only around 7%. Consequently, the ability to use the card payments data to predict total sales in services as published by the CZSO is limited.

An important methodological change regarding MC data reporting took place in 2016.[6] Therefore, we decided to use the MC data for the period March 2016–February 2022. Figures 1.a, 1.b, and 1.c visualize the month-on-month growth rates of sales in retail, accommodation, and restaurants – the

[4] Mastercard data is subject to controls related to data protection and regulations, which limits the broad use of the data. For practical use, Mastercard offers standardized platforms and products for accessing relevant insights and data (e.g., SpendingPulse™).

[5] Including cash withdrawals only increases the correlation between the MC data and the CZSO data for the category of all services (NACE 49–82) – see Section 4.1.

[6] Mastercard began enforcing reporting of transactions processed only by acquirers to Mastercard.

categories we focus on in this analysis. Because of increased month-on-month volatility during the covid-19 period, we split the sample to improve the readability of the figures: the left panel shows the data up to 2020 and the right shows the data from 2020 onward. It is obvious that in general the MC data fit the official CZSO data remarkably well.

Figure 1.a: Month-on-Month Growth Rates of Sales in Retail (NACE 47)

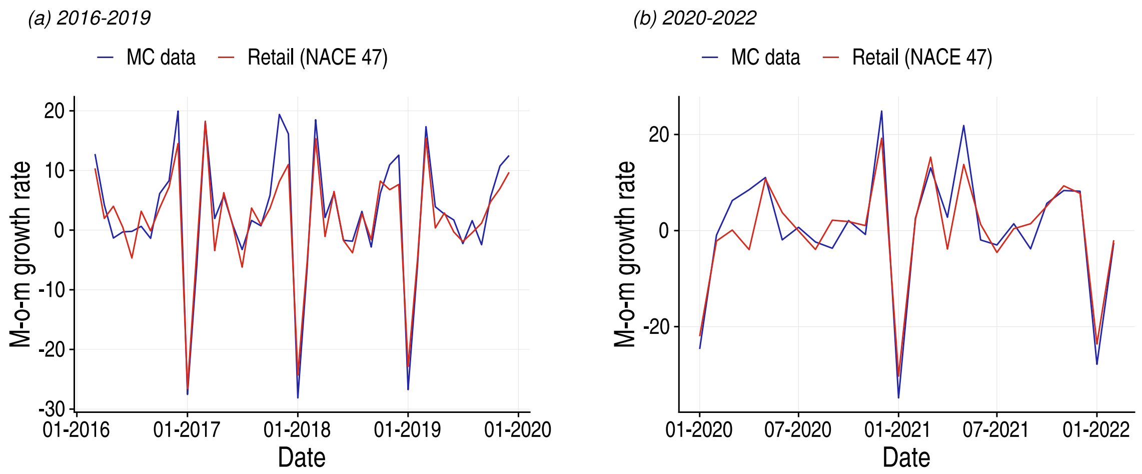


Figure 1.b: Month-on-Month Growth Rates of Sales in Accommodation (NACE 55)

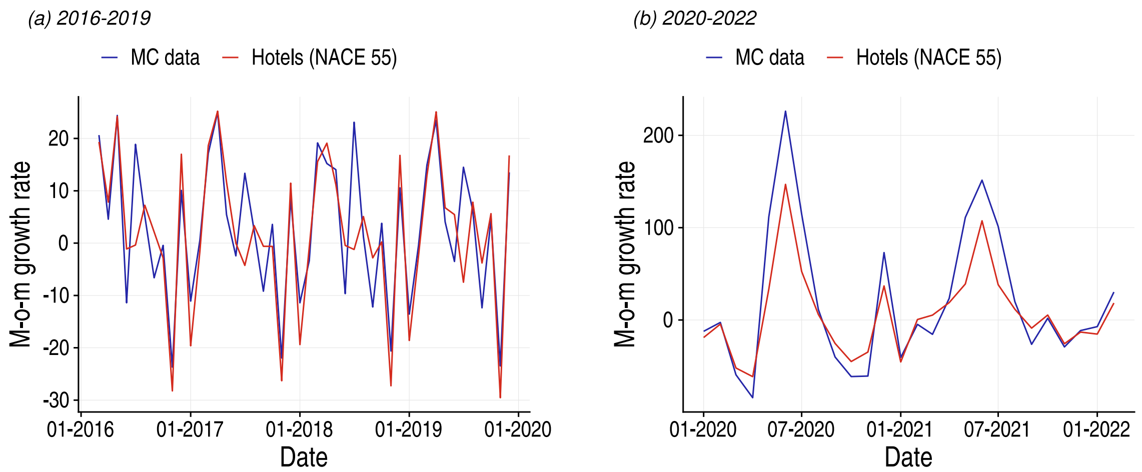
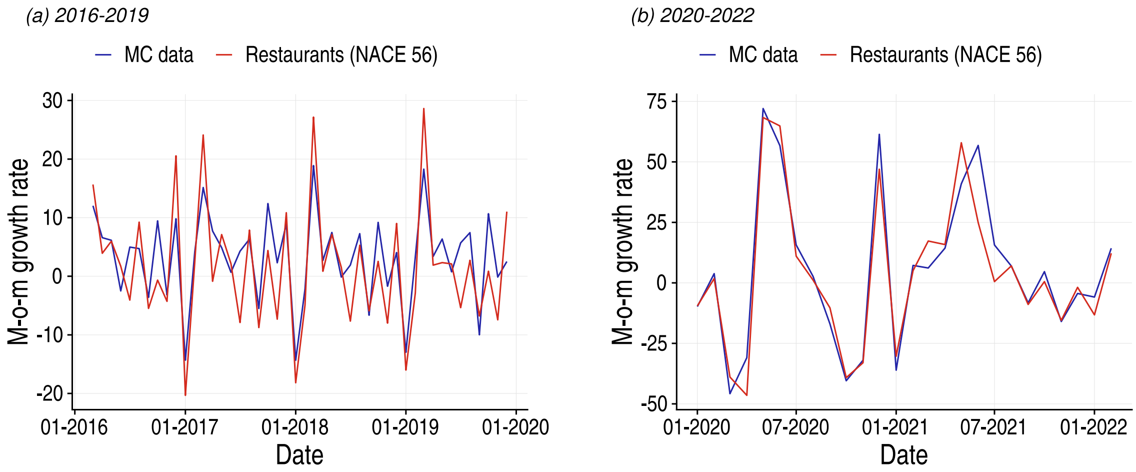


Figure 1.c: Month-on-Month Growth Rates of Sales in Restaurants (NACE 56)



2.3 Seasonal Adjustment of the Data

Figures 1.a–1.c clearly reveal strong seasonal patterns in both the CZSO and MC data. There are, in general, two ways of taking seasonality into account. The preferred way would be to first seasonally adjust the data and work with seasonally adjusted CZSO and MC time series. But as we mentioned above, the exact specification of the official adjustment method is not available. Figure 2 shows the replication of the adjustment of total retail sales using the X-13ARIMA-SEATS method, which reveals an important fact. It is clear that in the 2016–2019 period, the fit is rather low – the correlation coefficient is only 0.46. But during the covid-19 period, the adjustment method used by the CZSO visibly changed. The adjusted series becomes much more similar to the one obtained by the X-13 method, the correlation coefficient increasing to 0.93. Given the changing patterns of seasonality during the pandemic, the CZSO could probably no longer rely on its tailored methods and switched to a more mechanical approach. In line with this hypothesis is the fact that the CZSO practically abandoned any attempt to adjust the individual series of sales in accommodation and restaurants during the covid-19 pandemic (see Figures 3.a and 3.b).

The CZSO appears to have treated the different series differently when adjusting for seasonality during the pandemic. In our view, the inability to replicate the seasonal adjustment method used by the CZSO makes it impossible to consistently adjust the series for seasonality over the whole observed period. Attempting to do so would only introduce artificial noise into the data.

Therefore, we opt for the second approach: instead of adjusting the series prior to the analysis, we model seasonality during the backcasting and nowcasting exercises using the seasonal autoregressive integrated moving average (SARIMA) models or the seasonal random walk models. We then test whether the additional information resulting from the addition of unadjusted MC data improves the performance of the forecasting models which take into account the seasonality of the CZSO time series.

Figure 2: Comparison of Seasonal Adjustment of Total Retail Sales (NACE 47) by the CZSO and by the X-13 Method (m-o-m change in %)

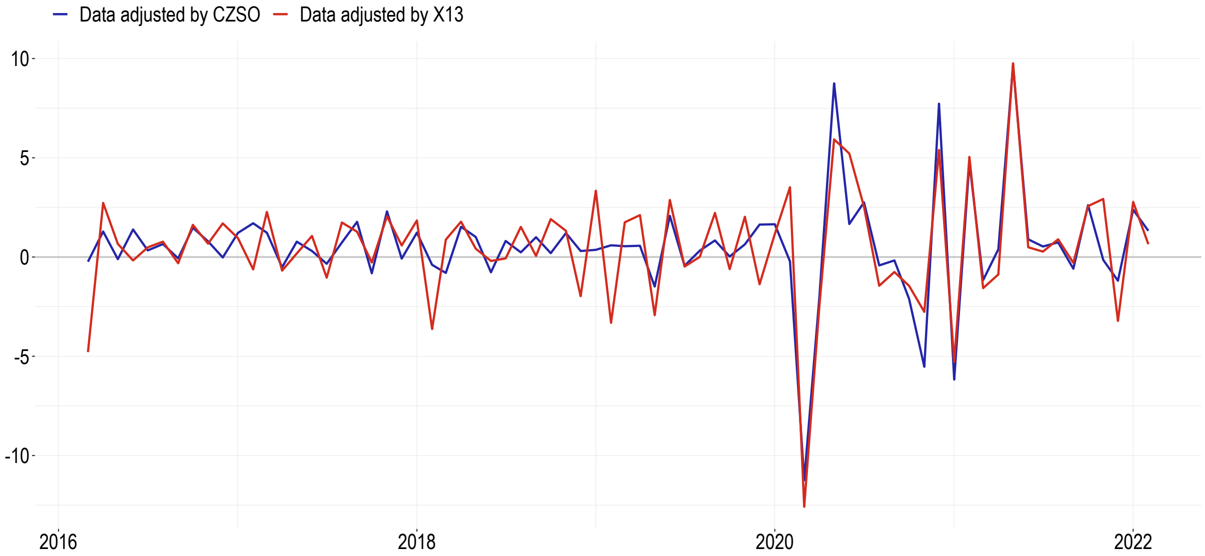


Figure 3.a: Sales in Accommodation (NACE 55) – Seasonally Adjusted and Unadjusted (m-o-m change in %)

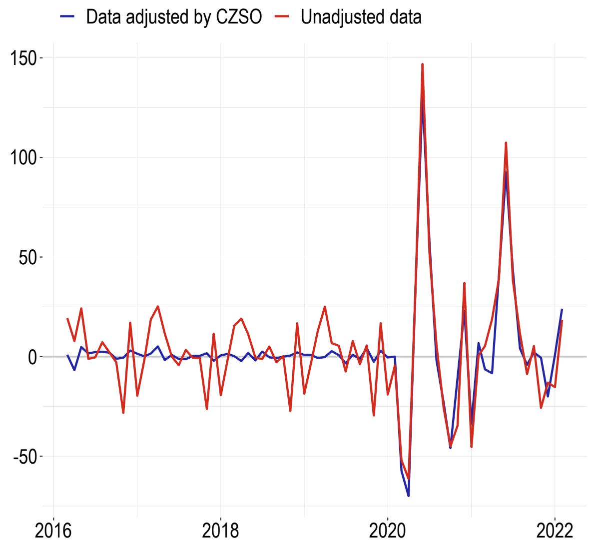
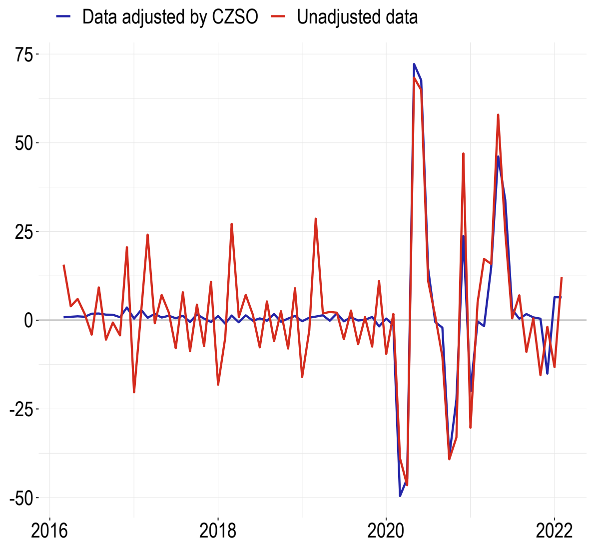


Figure 3.b: Sales in Restaurants (NACE 56) – Seasonally Adjusted and Unadjusted (m-o-m change in %)



3. Methodology

In order to assess the explanatory power of the MC data, we conduct comprehensive tests and nowcasting exercises. In the first step, we look at statistical significance tests of how the MC data fit the monthly indicators published by the CZSO, which we also refer to as “retail sales” further on. Next, we test the backcasting performance of the data, i.e., whether knowing the MC data for a whole given month is informative on retail sales volumes in that month. Finally, we conduct nowcasting exercises where we test whether knowing several daily observations of the MC data is informative about retail sales in the whole month.

One could ask why we do not aim to fit the MC data to the data on the consumption of households, which comprise about 50% of Czech GDP. We have several reasons for that. Card payments data in general do not contain most large purchases (e.g., car sales) or transactions paid typically by bank transfers (electricity and heating bills, for example). Next, it is clear that the MC data will be a better proxy for retail sales than for consumption as a whole. But even if we fitted retail sales perfectly, those series do not explain household consumption from the national accounts very well (see Appendix A). Finally, the sample of MC data is limited, and fitting quarterly household consumption data on a sample spanning less than 10 years could lead to imprecise conclusions.

As we explained above in Section 2.3, we were not able to reliably replicate the seasonal adjustment performed by the CZSO. Therefore, we use data that is neither seasonally nor calendar adjusted for the statistical tests. All data are in current prices. Our intention was to interfere with the data as little as possible so as not to affect the results.

3.1 Statistical Tests: How Well Can Mastercard Card Payments Data Explain Overall Retail Sales?

In the first step, we perform a correlation analysis between the MC data and the data published by the CZSO. We use MC data aggregated to monthly frequency and transformed to month-on-month growth rates. The correlation coefficient shows how tight the linear relationship between the two time series is. We test whether this coefficient is statistically significant or not.

Next, we consider a more complex relationship in the data by means of linear regressions. We regress monthly changes in retail sales (y_t) on monthly changes in the volume of card payments (x_t). Because of the strong seasonality, we consider three models:

(i) Simple model: $y_t = \alpha + \beta x_t + \epsilon_t$

(ii) $SARIMAX$

(iii) Restricted model: $y_t = x_t + \epsilon_t$

The first model explains changes in retail sales purely by changes in the volumes of card payments. The second model combines the seasonal autoregressive integrated moving average (SARIMA) model, whose order is chosen based on the Akaike information criterion, with the exogenous variable of card payments.

To assess the fit of the first model, we perform a series of standard econometric tests. We test whether coefficients α and β are significant by means of a t-test. We also test whether the regression as a whole is significant by means of an F-test. Finally, we test whether coefficient $\alpha = 0$ and coefficient $\beta = 1$ at the same time. Under this null hypothesis, knowing the raw MC data would be enough to estimate retail sales without any transformation and without the need to estimate any parameters (model (iii)). Models (ii) and (iii) are used only in assessing the backcasting and nowcasting performance of the MC data.

3.2 Backcasting Performance: Can We Approximate Retail Sales Using Mastercard Card Payments Data at the End of the Month?

In the next exercise, we assume that we know the MC data for a complete month, which is usually after the first week of the subsequent month. Our aim is to assess whether this data can approximate the time series published by the CZSO. Technically, this is a backcasting exercise, although the CZSO data is published with a significant delay (see Section 2.1). If this exercise confirms strong backcasting performance of the Mastercard data, it will be possible to accurately predict data four to five weeks before they are published.

To assess the backcasting performance of the data, we split the whole period for which data is available into a training sample (March 2016–December 2019) and a testing sample (January 2020–February 2022). We fit the model on the training sample and obtain estimates of the model parameters. Next, for each observation *t* in the testing sample, we estimate the change in retail sales ($\hat{y}_t$) by plugging in the known month-on-month change in the volume of card payments (x_t) and compute the forecast error. Finally, we compute two forecasting accuracy metrics: the root-mean-square error (RMSE) and the mean absolute error (MAE). In addition, we assess the probability with which the model estimated the correct sign of the change. Being able to correctly predict turning points in retail sales is of the utmost importance for monetary policy. We contrast these metrics with those obtained from two naïve benchmark models: the SARIMA model, with the order chosen based on the Akaike information criterion, and the seasonal random walk model. We also perform the Diebold-Mariano test to assess whether backcasts using transactional data are significantly better than the benchmark models.

3.3 Nowcasting Performance: Can We Approximate Retail Sales When We Know the Card Payments Data for the First Half of the Month?

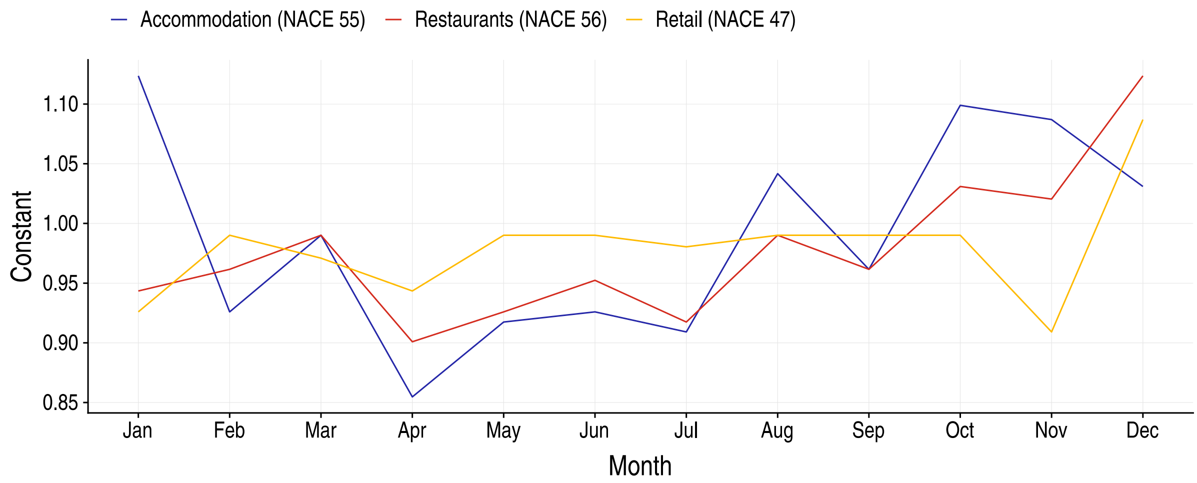
The nowcasting exercise at central banks usually involves estimating the quarterly growth rates of GDP, consumption, etc., using higher-frequency data, such as industrial production indices or sentiment indicators. Our aim in this paper is to derive the monthly growth rates of retail sales from the daily card transaction data.

Our approach resembles the bridge equations model (see Adam and Novotný, 2018, for example). As in the backcasting exercise, we split the data into a training sample and a testing sample. On the training sample, we fit models (i)–(iii).

In the testing period, we first forecast the whole-month MC data based on the first 14 days of a given month. This period was chosen so that we do not have to correct for seasonality inside a week, because the 14-day period always contains two weekends and 10 working days (for simplicity, we disregard national holidays). Then we calculate the ratios of the average daily sales in the first 14 days of each month to the average daily sales in the second half of each month. Figure 4 shows that the ratios are typically below 1, which means that average daily sales are typically larger in the second half of the month.

This might be due to the liquidity constraints on households, whose consumption increases after payday. In addition, the effect of larger sales in the second half of the month is particularly pronounced in November (when Christmas gifts are purchased) and January (due to the holidays at the beginning of the month) for total retail sales. On the other hand, December sales are significantly higher in the first half of the month due to purchases of Christmas gifts and due to the Christmas holidays.

Figure 4: Average Ratios of Mastercard Average Daily Sales in the First 14 Days of a Month vs. the Rest of the Month



In the next step, we take the sum of all card payments in the respective NACE category for the first 14 days of the month we are nowcasting and multiply it by the inverse of the ratio to get a whole-month MC data prediction. At the time the first 14 days of MC data are available, the CZSO data for the previous month are not yet published. Therefore, before estimating the retail sales nowcast, we need to first backcast the previous month.

To illustrate the exercise, let's assume we want to predict the sales in January after the MC data covering the first two weeks become available around the 21st of January. At that time, the last CZSO retail sales data available are for November. Therefore, depending on the model specification, we first estimate the December sales using either only MC data (models (i) and (iii)), or a combination of CZSO and MC data (model (ii)). Then we calculate a prediction for the whole-month January MC data as described above.

As the last step, we estimate the monthly change in retail sales for January and compute the forecast errors and forecasting accuracy metrics. We contrast them again with those obtained from the naïve benchmark models, only this time, due to the timing problem, the previous month has to be estimated by the SARIMA benchmark, too.

4. Results

Our aim is to use the MC data to nowcast the official Czech Statistical Office (CZSO) data on retail sales and sales in selected services. To fit the CZSO data, we construct NACE aggregates based on the Merchant Category Classification provided in the MC data. In this section, we first estimate the correlation between the MC data and the CZSO data to show how well the MC data track retail sales and sales in services. Then we assess the performance of the models used for backcasting the CZSO data using the MC data after they become available one week after the end of a month, which is four to five weeks before the CZSO data are released. And finally, we assess the ability to nowcast the CZSO data another two weeks sooner after the first two weeks of MC data become available.

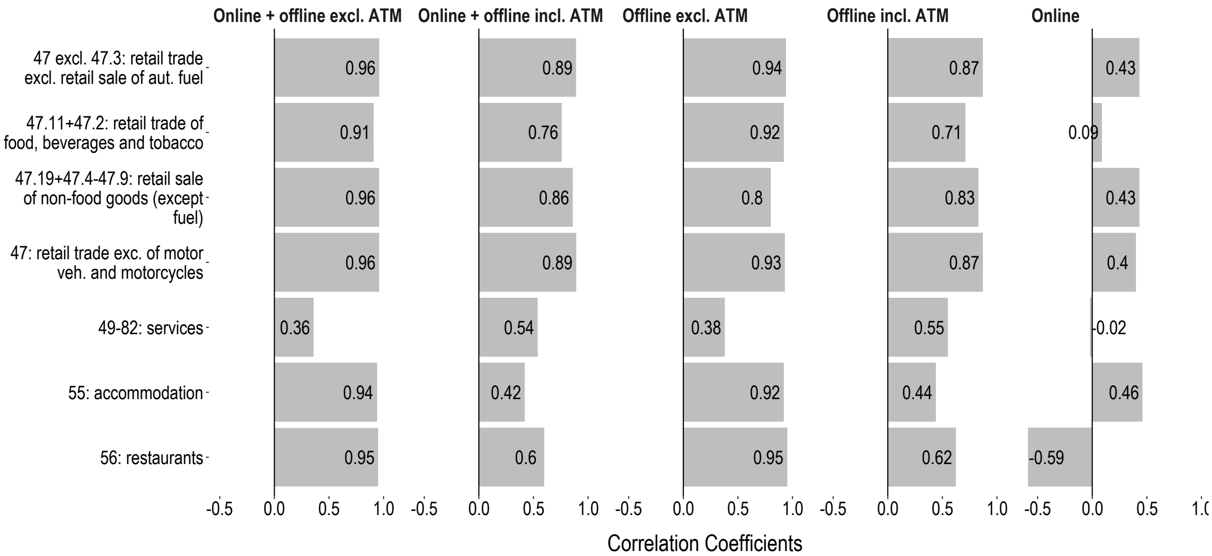
Most of the results are based on time series data from March 2016 to February 2022. RStudio was used for statistical analyses and data visualizations.[7]

4.1 Statistical Tests: The Card Payments Data Fits Total Retail Sales, Sales in Accommodation, and Sales in Restaurants Very Well

Correlation Analysis

We estimate the correlation coefficients between the different categories of card payments data and the data on retail sales and sales in services published by the CZSO. We consider the NACE categories described in Section 2.1. On the MC data side, we consider five sets for each category (the columns in Figure 5), depending on whether we included ATM withdrawals and online payments. The reason for including ATM withdrawals is that a significant proportion of transactions are still paid in cash, especially for services. As a result, ATM withdrawals could bear relevant information on consumption growth. The results of the correlation analysis can be found in Figure 5.

Figure 5: Pearson's Correlation Coefficients of Monthly Changes in Retail Sales and Transactional Data



Note: Online transactions consist of all transactions that are done in the *card not present* environment (i.e., e-shops). Offline transactions consist of all transactions in the *card present* environment (i.e., point of sales terminals)

Based on the correlation analysis, the MC data fit the CZSO data dynamics exceptionally well. We detected the closest correlation for the most general category of retail, i.e., *retail trade, except of motor vehicles and motorcycles* (NACE 47). Pearson's correlation coefficient was estimated at 0.96, which suggests very tight correlation. The same result was obtained for *retail trade excl. retail sale*

[7] RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

of automotive fuel (NACE 47 excl. 47.3) and *retail sale of non-food goods except fuel* (47.19+47.4–47.9). A high correlation coefficient (0.91) was estimated for *retail trade of food, beverages and tobacco* (NACE 47.11+47.2).

The fit of the transactional data is weaker for the turnover in services. The correlation coefficient in the aggregated services group (NACE 49–82) was estimated at 0.36, indicating a lower to medium level of correlation between the MC data and the CZSO series. The reason for this might be obvious, as electronic card payments are not present in most of the subcategories. In the case of the accommodation and restaurants subgroups, which constitute a large proportion of the card payments data, the correlation coefficients are high, with values of 0.94 and 0.95, respectively.

Regarding ATM withdrawals, the correlation coefficients generally weaken when we include data on cash withdrawals. A notable exception is the services category, where including data about ATM withdrawals improved the correlation from 0.36 to 0.54. In the case of services, card transactions constitute only a relatively small proportion of the CZSO data (see Figure C2 in Appendix C), which indicates the reason for the low level of correlation. Additional information about ATM withdrawals hence improved the correlation of these series, not necessarily indicating that cash is being used in these types of units.

Regarding e-commerce, including online stores in the MC data generally increases the correlation coefficients and results in a closer relationship with the CZSO series. The exceptions are the categories of retail trade of food, beverages and tobacco, services, accommodation, and restaurants.

Based on the correlation analysis, we identified the following three categories for our subsequent analysis: retail (NACE 47), accommodation (NACE 55), and restaurants (NACE 56). We consider card payments data which are not complemented by ATM withdrawals. These series have the strongest correlation across the tested groups and indicate strong potential for the use of card payments data in the nowcasting exercise.

Linear Regressions

Next, we test more complex relationships in the data using linear regression models (models (i) and (iii)). The results are summarized in Table 1. Additional results can be found in Table B1 in Appendix B.

Similarly to the correlation analysis, the fit of all the models is very strong. The fit is best for the retail category, as suggested by the coefficient of determination, which is higher than 0.9. For turnover in accommodation and restaurants the fit is slightly lower, but R^2 is still significant and higher than 0.85. Regarding the regression coefficients, only coefficient β (MC data) is significant for all categories and models, and it is closest to one in the restaurants category.

Interestingly, the restricted model fits the data well only in the restaurants category. In the other categories, the fit of the restricted model is significantly worse than the fit of the unrestricted model. Since all coefficients β are smaller than 1, the transformation of the MC data reduces its volatility. But being able to better fit the data does not necessarily imply better nowcasting ability. Therefore, we assess the precision of the restricted model predictions as well. The following sections show

how the forecast errors change when one uses the restricted model compared to the unrestricted ones.

Table 1: Summary of Regression Models

Category	Model	R^2	α	β	Does restricted model fit?
Retail	(i)	0.91	−0.44 (0.35)	0.81*** (0.03)	✗
Accommodation	(i)	0.88	−1.74 (1.27)	0.60*** (0.03)	✗
Restaurants	(i)	0.86	−1.46 (0.89)	0.94*** (0.04)	✓

Note: (i) $y_t = \alpha + \beta x_t + \epsilon_t$; */**/*** represent significance levels of 0.1/0.05/0.01
Restricted model: model (iii) $y_t = x_t + \epsilon_t$

4.2 Backcasting Performance: The Card Payments Data Significantly Improves the Ability to Predict the CZSO Data

In this section, we address the ability of the whole-month MC data, which become available one week after the end of a month, to improve the prediction of the CZSO data published another four to five weeks later. The backcasting performance of models (i)–(iii) is summarized in Figure 6 and the detailed results are presented in Tables B2 – B4. In the backcasting exercise, the models were trained on the period between March 2016 and December 2019. The models were subsequently tested on the period between January 2020 and February 2022.

We measure the performance of the models by comparing the predicted values with the values observed and published by the CZSO after all revisions. We use three metrics: the root-mean-square error (RMSE), the mean absolute error (MAE), and the sign error (which measures the proportion of incorrect predictions of the direction of the monthly change in sales – either an increase or a decline). As a benchmark, we consider two models, first, a seasonal random walk model, where the prediction for a specific month is the month-on-month change for the same month one year before, and second, a SARIMA model estimated solely on the CZSO data between February 2000 and December 2019 and tested on the period January 2020–February 2022. We decided to use the longest possible training sample, because we did not want to penalize the CZSO data due to the limited availability of the MC data.[8]

Another important caveat regarding the CZSO data is that our analysis uses the final retail sales data after all revisions. As we mention in Section 2.1, the data is first revised and corrected together with the publication of the data for the following month but is also subject to a final revision of the whole-year monthly data after the publication of the December sales, usually in March of the following year. Because the real-time data vintages of retail sales and sales in services are not available from the CZSO in the required form, we assess the ability of our models to predict the

[8] We performed the usual diagnostics for the SARIMA benchmark and the SARIMAX model (ii). The residuals show no obvious patterns, the ACF and the PACF do not reveal any remaining correlations, and the Ljung-Box test does not reject the null hypothesis of independent and identically distributed residuals. The test results are available upon request.

final revised retail sales, which are, in fact, available up to 13 months later. The revisions can be significant, particularly during periods of heightened economic volatility and uncertainty, such as during the covid-19 period. In 2020, the revisions of the month-on-month growth rates of total retail sales ranged from −2.3 percentage points in December to +2.2 pp in April. The average revision was 1.3 pp. In 2021, the revisions still ranged from −1.6 pp to +2.4 pp, with an average revision of 0.9 pp.[9] As a consequence, we test the ability of the MC data to nowcast and backcast the fundamental evolution of retail sales, and not necessarily how the CZSO captures them in real time.

In general, the MC data perform well in the backcasting exercise. The RMSE and MAE metrics drop by about 40–70% compared to the best benchmark when we use the MC data. Looking at the MAE values, the models perform reasonably well: the MAE of the best-performing prediction model is 30% of the mean of the absolute-value month-on-month changes over the testing period for total retail sales, 39% for accommodation, and 29% for restaurants.

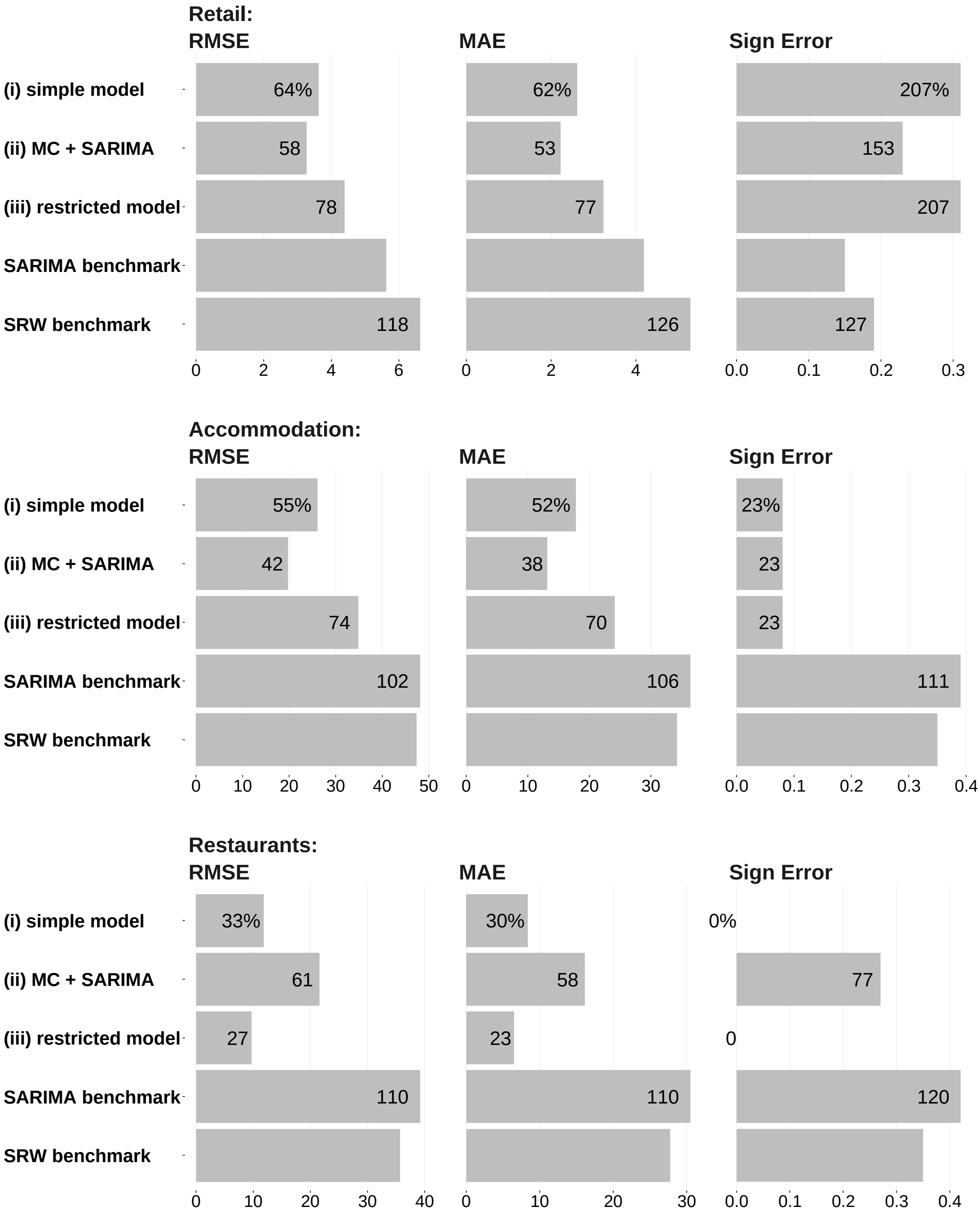
As we expected, including seasonal terms (model (ii)) improves the fit of the models explaining retail sales and sales in accommodation. In the case of sales in restaurants, the restricted model (iii) performs best for the chosen testing period January 2020–February 2022. This is consistent with the findings in Section 2.3 and further confirms that the covid-19 lockdowns in all probability completely changed the historical seasonal patterns for this type of service.

Regarding the sign errors, their proportion is higher for retail sales as a whole but lower for accommodation and restaurants compared to the benchmarks. Still, in more than 70% of observations, the direction of the change in sales is correctly predicted for the retail sales aggregate.

To sum up the results, the forecast errors are lower when we consider the data from MC compared to the benchmarks. At the same time, the sign of the changes in sales is predicted slightly worse, but even this metric is still satisfactory. The Diebold-Mariano test confirms the higher predictive accuracy of the tested model (ii) for all three categories compared to the best-performing benchmark models.

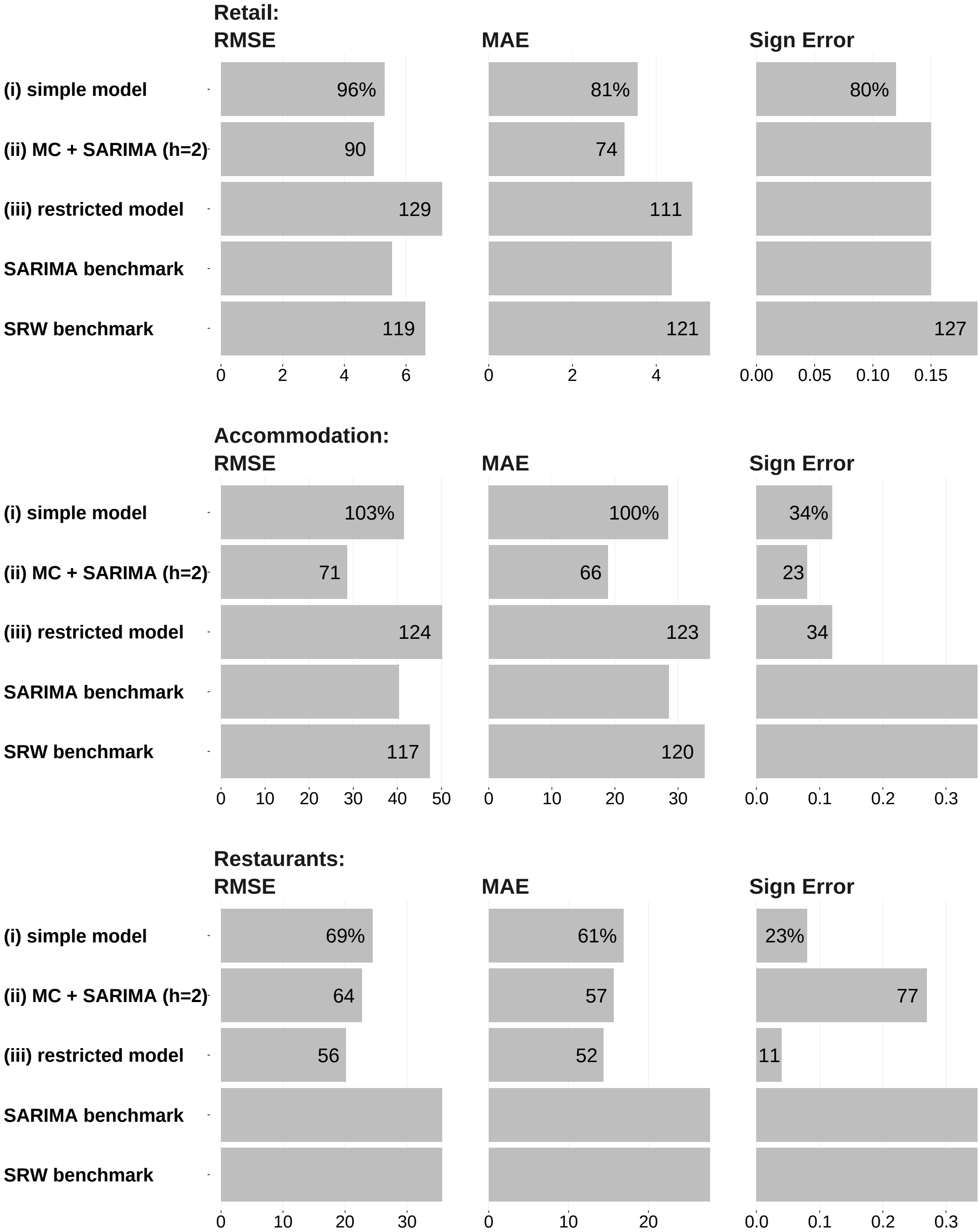
[9] Based on the comparison of seasonally and calendar unadjusted sales in constant prices.

Figure 6: Backcasting Performance



Note: SRW benchmark stands for the seasonal random walk model. The percentages represent the size of the respective error relative to the best-performing benchmark.

Figure 7: Nowcasting Performance



Note: SRW benchmark stands for the seasonal random walk model. The percentages represent the size of the respective error relative to the best-performing benchmark.

4.3 Nowcasting Performance: Knowing the Card Payments Data for the First Half of a Month Still Trumps the Benchmark

We can improve the predictive accuracy compared to the benchmark model if we use the full-month MC data. But what if we wanted to estimate retail sales in a month by using data covering only the first 14 days of this month? This forces us to estimate not only the end-of-month sales based on the MC data, but also the previous month's result, which is still not published at that time. In other words, this exercise addresses the trade-off between accuracy and timeliness.

In terms of the training and testing sample and the performance metrics, the nowcasting exercise is structured in the same way as the backcast. The results are summarized in Figure 7 and described in detail in Tables B5 – B7.

Overall, using card payments data covering only the first 14 days of the month in question does not worsen the predictive accuracy compared to the benchmark models. And as we show below, the added information turns out to be especially valuable in periods of abrupt economic changes, such as the introduction and lifting of lockdowns and other measures during the covid-19 pandemic.

Just like in the case of the backcast, the MC data improves the prediction accuracy the most for the category of sales in restaurants – by more than 40% compared to the best-performing benchmark. But the accuracy also increases for total retail sales (by 10%) and accommodation (by 29%). In addition, the sign errors are decreased in the accommodation and restaurants category and unchanged for total retail sales. As expected, the absolute performance is worse compared to the backcasts: the ratio of the MAE to the mean of the absolute-value month-on-month changes over the testing period increases to 44% for retail sales, 56% for accommodation, and 64% for restaurants. Looking at the individual models for each category, model (ii), i.e., the model including seasonality and the MC data, performs very well and improves the average accuracy compared to the best-performing benchmark model for all categories, although in the case of restaurants, the restricted model (iii) again provides the highest prediction accuracy.

To sum up the results, knowing 14 days of transactional data tells us a lot about the direction of the change in sales in a given month, i.e., whether the sales will increase or decrease. Even with the highest sign error for the total retail sales category, we are able to correctly predict the direction of the change in 85% of cases. And even though the absolute performance of the models is worse than when knowing the end-of-month values, the models with MC data work better than both benchmarks and generate predictions with reasonable error metrics.

4.3.1 Nowcasting Performance during Covid-19: Better Than Expected

The practical implications of the nowcasting performance of the MC data can be illustrated on the covid-19 period, which can be characterized by rapid changes in the economic environment. We focus on three periods: the first pandemic wave in March–May 2020, the second wave in October–November 2020, and the reopening of the economy in May–June 2021. The choice of the models used for the calculations in this section is based on their backcasting and nowcasting performance: model (ii) for retail sales and sales in accommodation, and model (iii) for sales in restaurants. But the important trend changes in the periods described can be deduced using any of the estimated models, including the naïve model (iii), which assumes that the change in sales perfectly reflects the change in card payments. As benchmarks, we always use the best-performing benchmarks for backcasts: SARIMA for retail sales, and seasonal random walk for sales in accommodation and restaurants.

The Czech government's first response to the pandemic was to declare a state of emergency on March 12, 2020. Two days later, most retail sales and sales of services in business premises were forbidden. Some businesses started to reopen during April. The rest, including restaurants and accommodation services, were allowed to reopen only by the end of May.

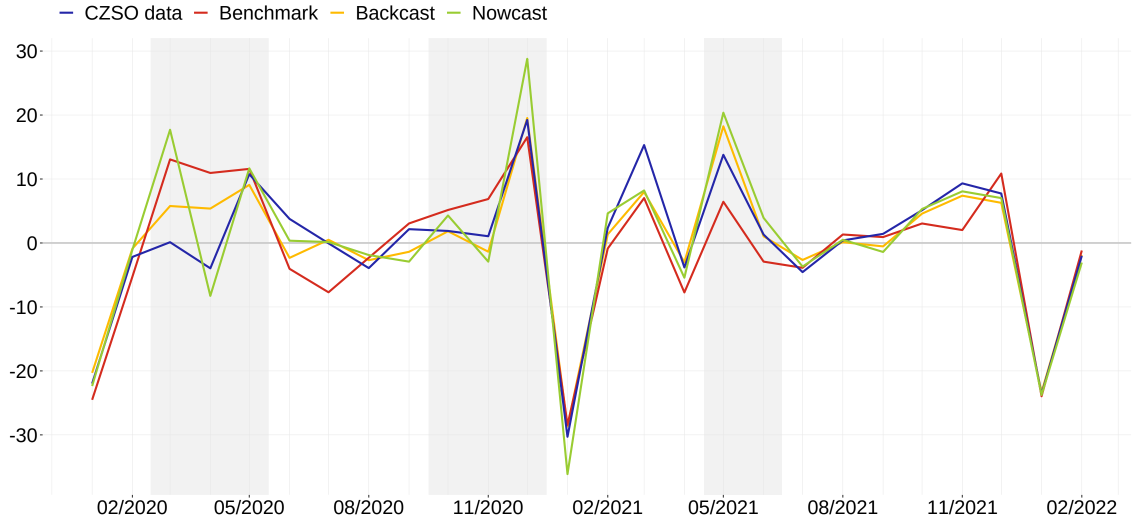
Relying solely on the CZSO data, the first partial quantification of the impact on household consumption in March became available on May 6. The data for April, the first whole month of the covid-19 pandemic, was published on June 5, 2020. Predictions based on the benchmark SARIMA model utilizing only the CZSO data were not able to capture the break in the time series driven by strong seasonality (see Figures 8–10). But the MC data were able to reveal a significant slowdown already after the first week in April – the backcast of the March growth in retail sales (5.8%) was still higher than the actual figure (0.1%) but pronouncedly lower than the benchmark prediction (13.1%). The predicted fall in sales in accommodation (−42.8%) and restaurants (−45.8%) was remarkably close to the official numbers (−52% and −38.9%) published much later, whereas the benchmark predicted growth of 12.9% in accommodation and 28.6% in restaurants.

Another two weeks later, based on the first 14 days of the MC data, the nowcast model predicted an 8.3% month-on-month decline in retail sales in April (−4% in reality). The gradual move of a large proportion of retail online and a consequent increase in the share of card payments biased the whole-month estimate (backcast) for April slightly upward, but it was still more than 5 percentage points below the benchmark prediction and closer to the CZSO data. After the first 14 days of May, the nowcast accurately predicted a strong rebound in sales due to the reopening of the economy.

The second major wave of covid-19 broke out in October 2020 and led to a complete lockdown by the end of October, including again the prohibition of almost all retail sales and the sale and provision of services in establishments. In contrast to the usual seasonality, the MC data predicted a significant slowdown of sales in November and a rebound in December following the reopening of the economy on December 3. The December backcast of retail sales almost perfectly fits the official CZSO data published in February.

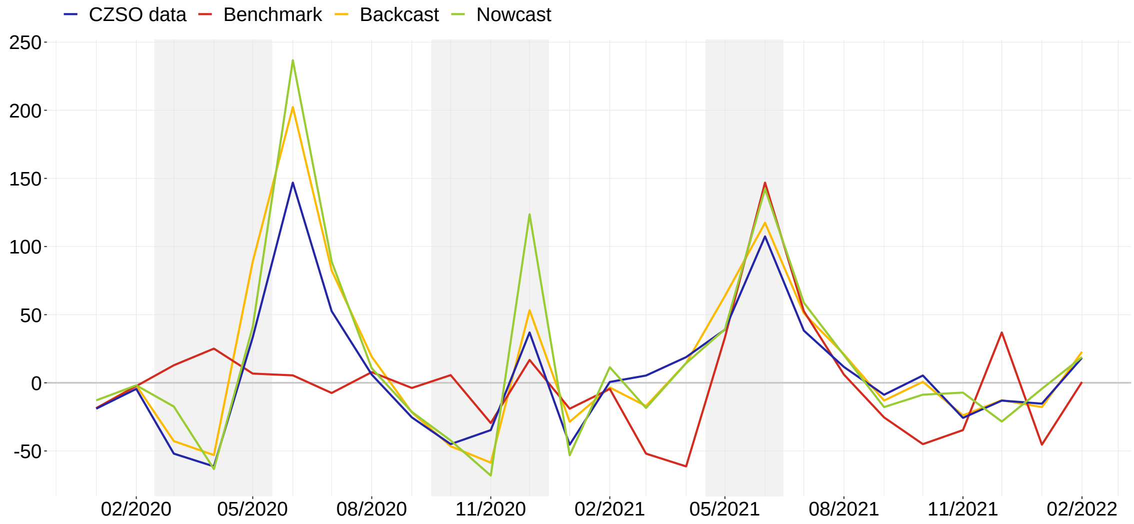
At the end of December, almost all shops and service providers were closed again. The anti-pandemic measures were then repeatedly relaxed and tightened in reaction to the evolution of the pandemic situation. This continued until May 2021. The MC data again confirmed the revival of demand in timely fashion. But, even more importantly, the June data revealed that the upsurge in retail sales was only temporary as an effect of deferred consumption. The same was observed in June–August for sales in restaurants, which opened at the very end of May. The ability to see almost in real time that household demand for goods and especially services stayed rather subdued would have been a crucial piece of information for monetary policy makers during the second half of 2021.

Figure 8: Retail Sales (CZSO data, benchmark, backcast, nowcast)



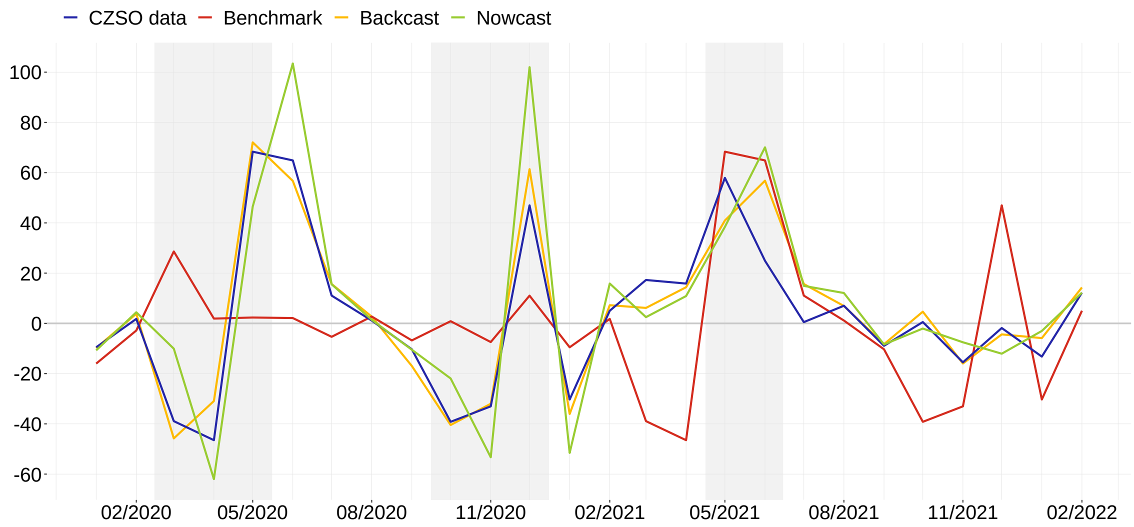
Note: SARIMA backcast used as the benchmark.

Figure 9: Sales in Accommodation (CZSO data, benchmark, backcast, nowcast)



Note: Seasonal random walk used as the benchmark.

Figure 10: Sales in Restaurants (CZSO data, benchmark, backcast, nowcast)



Note: Seasonal random walk used as the benchmark.

4.3.2 Backcasting Performance before Covid-19: The MC Data Doesn't Outperform the Benchmark

The backcasting and nowcasting exercises were performed on a relatively short sample, and the testing period was very specific in that it contained the period of covid-19. Even though we believe that the timing advantage[4] is especially useful during times of economic distress, in this section we check whether the availability of whole-month MC data lets us improve the prediction of retail sales and sales in accommodation and services.

To test the explanatory power of the card payments data in normal economic times, we perform a backcasting exercise using the restricted model ($y_t = x_t$, where y_t is again the monthly change in sales in the respective category and x_t is the monthly change in the volume of card payments) on the period between 2016 and 2019. Since we use the restricted model, we do not have to estimate any parameters in the card payments data model. Therefore, we do not need to split the sample into training and testing periods.

Table 2 shows that before covid-19, the MC data do not improve the prediction compared to the seasonal random walk benchmark model for any of the categories analyzed. Moreover, for sales in accommodation services and restaurants, the restricted model performs significantly worse than the benchmark. The reason is the strong seasonality present in the data in a normal macroeconomic situation (see Figures 1.a–1.c). This further proves that the primary advantage of using the MC data lies in the ability to backcast and nowcast retail sales in situation where seasonal patterns break down, as in the case of the covid-19 pandemic, as we showed in Section 2.3.

But still, looking at the MAE values, the MC data predict the CZSO data reasonably well, especially for total retail sales: over the analyzed period of March 2016–December 2019, the MAE of the restricted model is 38% of the mean of the absolute-value month-on-month changes for total retail sales, 54% for accommodation, and 64% for restaurants.

Table 2: Backcasting Performance before Covid-19

	Model	RMSE	MAE	Sign error	Is it worse? [1]
Retail	(iii) restricted model	3.21	2.47	15%	No
	SARIMA benchmark	5.55	2.95	20%	
	SRW benchmark	3.08	2.49	22%	
Accommodation	(iii) restricted model	7.95	5.81	20%	Yes**
	SARIMA benchmark	7.00	4.43	14%	
	SRW benchmark	4.73	3.40	15%	
Restaurants	(iii) restricted model	6.14	5.07	22%	Yes***
	SARIMA benchmark	4.64	3.38	17%	
	SRW benchmark	3.63	2.65	11%	

Note: */**/*** represent significance levels of 0.1/0.05/0.01; SRW benchmark stands for the seasonal random walk model; [1] The Diebold-Mariano test shows whether the restricted model is less accurate than the best-performing benchmark model (SRW for all three categories).

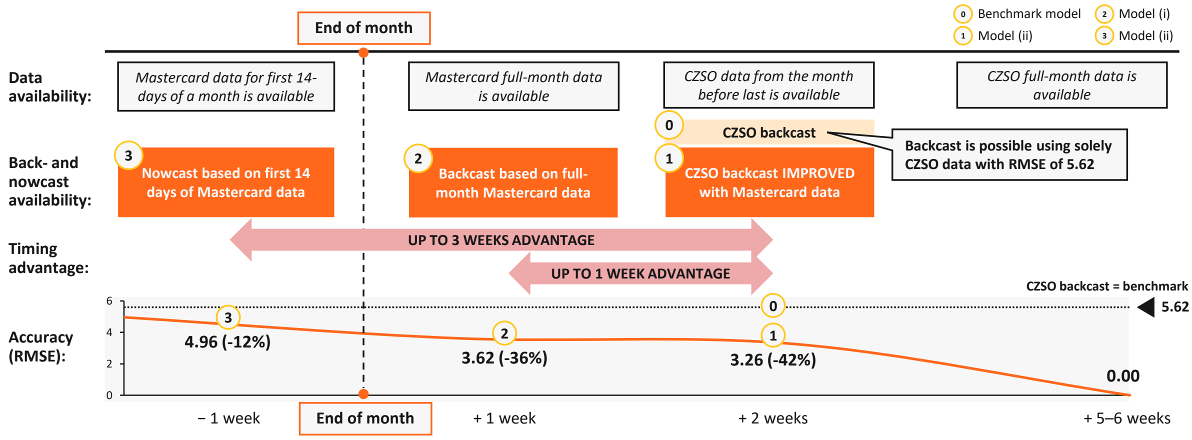
4.4 Implications for the Practical Use of Card Payments Data in Retail Sales

The card payments data are available at high frequency and significantly sooner than the CZSO data.[4] When backcasting the end of the current month total retail sales, we can use either a seasonal random walk model with significantly lower accuracy with no lag, or a SARIMA process based on the CZSO data only, which we call the "CZSO backcast" for simplicity. The CZSO backcast is available one to two weeks after the end of the backcasted month with an RMSE of 5.62 and constitutes our benchmark model. The addition of MC data into the CZSO backcast improves the prediction accuracy by 42%. When we use solely card payments data, backcasting of current month retail sales is available up to one week sooner and with 36% improved accuracy compared to the CZSO backcast. This is possible due to the timely availability of MC data for the backcasted month (see the results for model (i)).

However, the predictions can be available even sooner, thanks to the high frequency of the card payments data. A model utilizing MC data covering the first 14 days of a month still slightly improves the accuracy of the predictions (by 12%) and can nowcast current month retail sales up to three weeks sooner and six to seven weeks before the full-month data is published by the CZSO. Figure 11 summarizes these results and emphasizes the inherent trade-off between timeliness and accuracy.

The presented results prove that the accuracy of the predictions of current month sales varies considerably with the amount of information available to the analyst or decision-maker at the given moment. In this context, card payments data, with its timely availability[4] and high frequency, is an important contributor to nowcasts.

Figure 11: Timeliness vs. Accuracy Trade-off for Total Retail Sales



5. Conclusion

Makers of economic policy, including monetary policy, need timely information about current developments in the economy. Retail sales, published at monthly frequency with a delay of five to six weeks, are one of the most important sources of such information. But as the covid-19 pandemic showed us, such a lag may be too long in situations where the economic environment is changing rapidly.

In this paper, we use Mastercard card payments data ("MC data") to nowcast turnover in Czech retail sales and services. We show that an index based on this data tracks surprisingly well the official retail sales data released by the Czech Statistical Office ("CZSO data") more than a month later. We further show that the MC data not only helps in backcasting Czech retail sales after the end of the month (when the payments data is complete), but also provides valuable information for the nowcast as soon as three weeks into the ongoing month. This is six to seven weeks ahead of the official CZSO release.

We compare the performance of our backcasting and nowcasting models with two sets of benchmarks: the seasonal autoregressive integrated moving average (SARIMA) models and the seasonal random walk models. To assess the performance, we split the whole period for which data is available into a training sample (March 2016–December 2019) and a testing sample (January 2020–February 2022) and focus on three categories of retail and services: total retail (NACE 47), accommodation (NACE 55), and restaurants (NACE 56). For all three categories, the use of MC data significantly improves the forecasting performance over the benchmarks used. This is not the case for the pre-covid-19 period 2016–2019, where the seasonal random walk model performs better, partly due to the limited availability of MC data.

To illustrate the usefulness of our method, we show that we would have been able to backcast, with reasonable accuracy, the sharp drop in retail sales that occurred at the outbreak of the first wave of covid-19 in Czechia in the second half of March 2020 four weeks before the CZSO data release. This underscores the practicality and usefulness of our method in capturing significant shifts in the economic landscape.

References

AASTVEIT, K. A., T. M. FASTBO, E. GRANZIERA, K. S. PAULSEN, AND K. N. TORSTENSEN (2020): "Nowcasting Norwegian Household Consumption with Debit Card Transaction Data." Norges Bank Research Working Paper No. 17/2020.

ADAM, T. AND F. NOVOTNÝ (2018): "Assessing the External Demand of the Czech Economy: Nowcasting Foreign GDP Using Bridge Equations." Working Papers 2018/18, Czech National Bank.

ADAM, T., O. MICHÁLEK, A. MICHL, AND E. SLEZÁKOVÁ (2021): "The Rushin Index: A Weekly Indicator of Czech Economic Activity." Working Papers 2021/4, Czech National Bank.

ALCEDO, J., A. CAVALLO, B. DWYER, P. MISHRA, AND A. SPILIMBERGO (2022): "E-Commerce during COVID: Stylized Facts from 47 Economies." IMF Working Paper No. 22/19.

AMBRIŠKO, R. (2022): "Nowcasting Macroeconomic Variables Using High-Frequency Fiscal Data." Working Papers 2022/05, Czech National Bank.

ARNOŠTOVÁ, K., D. HAVRLANT, L. RŮŽIČKA, AND P. TÓTH (2011): "Short-Term Forecasting of Czech Quarterly GDP Using Monthly Indicators." *Czech Journal of Economics and Finance (Finance a úvěr)* 61(6), pp. 566–583.

BARLAS, A. B., S. G. MERT, B. O. ISA, A. ORTIZ, T. RODRIGO, B. SOYBILGEN, AND E. YAZGAN (2021): "Big Data Information and Nowcasting: Consumption and Investment from Bank Transactions in Turkey." arXiv preprint arXiv:2107.03299.

CHAPMAN, J. T. AND A. DESAI (2021): "Using Payments Data to Nowcast Macroeconomic Variables during the Onset of COVID-19." Bank of Canada Staff Working Paper 2021-2.

FRANTA, M., D. HAVRLANT, AND M. RUSNÁK (2016): "Forecasting Czech GDP Using Mixed-Frequency Data Models." *Journal of Business Cycle Research* 12(2), pp. 165–185.

GALBRAITH, J. W. AND G. TKACZ (2018): "Nowcasting with Payments System Data." *International Journal of Forecasting* 34(2), pp. 366–376.

GARCÍA, J. R., M. PACCE, T. RODRIGO, P. R. DE AGUIRRE, AND C. A. ULLOA (2021): "Measuring and Forecasting Retail Trade in Real Time Using Card Transactional Data." *International Journal of Forecasting* 37(3), pp. 1235–1246.

RUSNÁK, M. (2016): "Nowcasting Czech GDP in Real Time." *Economic Modelling* 54(C), pp. 26–39.

VERBAAN, R., W. BOLT, AND C. VAN DER CRUIJSEN (2017): "Using Debit Card Payments Data for Nowcasting Dutch Household Consumption." DNB Working Paper, 571.

Appendix A: Monthly Indicators of Demand

This appendix shows how monthly indicators of household demand (the volume of retail sales, turnover in services) are related to the consumption of households, which is a part of the quarterly national accounts and comprises about 50% of GDP.

We aggregate the monthly retail sales indicators to quarterly frequency and contrast the quarterly growth rate of the resulting time series with consumption by calculating correlation coefficients (Table A1). Even though the level of correlation of some categories is relatively high, the ability to match the variation in household consumption is rather limited for the aggregates of total sales in retail and services.

The implication is that even if we fitted the monthly indicators perfectly using the transactional data, we would still have only an imprecise estimate of the consumption aggregate.

Table A1: Correlations of Monthly Indicators with Consumption

	constant prices, seasonally adjusted	**current prices, not seasonally adjusted**
Retail – total	0.75	0.81
Retail trade and repair of motor vehicles and motorcycles	0.59	0.61
Retail trade, except of motor vehicles and motorcycles	0.71	0.86
Services – total	0.76	0.87
Accommodation and restaurants	0.87	0.69

Appendix B: Results

This appendix provides detailed results explained further in Section 4.

Table B1: Statistical Tests for the Restricted Model $y_t = x_t + \epsilon_t$

Category	Null hypothesis	F statistic	P-value (F-stat)
Retail	$\alpha = 0, \beta = 1$	22.44	0.00***
Accommodation	$\alpha = 0, \beta = 1$	118.54	0.00***
Restaurants	$\alpha = 0, \beta = 1$	2.70	0.07

Note: */**/*** represent significance levels of 0.1/0.05/0.01

Table B2: Estimated Errors of Backcasting Performance for Retail Sales

Model	RMSE	MAE	Sign error	P-value D-M test
(i)	3.62	2.61	31%	0.01***
(ii)	3.26	2.22	23%	0.00***
(iii)	4.39	3.23	31%	0.05**
ARIMA(2,0,0)(0,1,2)[12]	5.62	4.18	15%	
Seasonal random walk	6.62	5.28	19%	

Note: */**/*** represent significance levels of 0.1/0.05/0.01; the Diebold-Mariano test uses the alternative hypothesis “greater,” meaning the tested models (i)–(iii) are more accurate than the best-performing benchmark model.

Table B3: Estimated Errors of Backcasting Performance for Sales in Accommodation

Model	RMSE	MAE	Sign error	P-value D-M test
(i)	26.10	17.81	8%	0.03**
(ii)	19.71	13.15	8%	0.01***
(iii)	34.88	24.07	8%	0.09*
ARIMA(0,0,2)(0,1,1)[12]	48.09	36.38	39%	
Seasonal random walk	47.37	34.19	35%	

Note: */**/*** represent significance levels of 0.1/0.05/0.01; the Diebold-Mariano test uses the alternative hypothesis “greater,” meaning the tested models (i)–(iii) are more accurate than the best-performing benchmark model.

Table B4: Estimated Errors of Backcasting Performance for Sales in Restaurants

Model	RMSE	MAE	Sign error	P-value D-M test
(i)	11.84	8.37	0%	0.00***
(ii)	21.58	16.10	27%	0.00***
(iii)	9.67	6.47	0%	0.00***
ARIMA(0,0,1)(0,1,0)[12]	39.16	30.47	42%	
Seasonal random walk	35.66	27.70	35%	

Note: */**/*** represent significance levels of 0.1/0.05/0.01; the Diebold-Mariano test uses the alternative hypothesis “greater,” meaning the tested models (i)–(iii) are more accurate than the best-performing benchmark model.

Table B5: Estimated Errors of Nowcasting Performance for Retail Sales

Model	RMSE	MAE	Sign error	P-value (D-M test)
(i)	5.30	3.55	12%	0.23
(ii)	4.96	3.24	15%	0.02**
(iii)	7.17	4.86	15%	0.90
ARIMA(2,0,0)(0,1,2)[12]	5.54	4.37	15%	
Seasonal random walk	6.62	5.28	19%	

Note: */**/*** represent significance levels of 0.1/0.05/0.01; the Diebold-Mariano test uses the alternative hypothesis "greater," meaning the tested models (i)–(iii) are more accurate than the best-performing benchmark model.

Table B6: Estimated Errors of Nowcasting Performance for Sales in Accommodation

Model	RMSE	MAE	Sign error	P-value (D-M test)
(i)	41.47	28.43	12%	0.56
(ii)	28.63	18.86	8%	0.03**
(iii)	50.15	35.04	12%	0.92
ARIMA(0,0,2)(0,1,1)[12]	40.39	28.53	35%	
Seasonal random walk	47.37	34.19	35%	

Note: */**/*** represent significance levels of 0.1/0.05/0.01; the Diebold-Mariano test uses the alternative hypothesis "greater," meaning the tested models (i)–(iii) are more accurate than the best-performing benchmark model.

Table B7: Estimated Errors of Nowcasting Performance for Sales in Restaurants

Model	RMSE	MAE	Sign error	P-value (D-M test)
(i)	24.46	16.90	8%	0.03**
(ii)	22.73	15.66	27%	0.00***
(iii)	20.14	14.37	4%	0.00***
ARIMA(0,0,1)(0,1,0)[12]	35.65	27.70	35%	
Seasonal random walk	35.66	27.70	35%	

Note: */**/*** represent significance levels of 0.1/0.05/0.01; the Diebold-Mariano test uses the alternative hypothesis "greater," meaning the tested models (i)–(iii) are more accurate than the best-performing benchmark model.

Appendix C: Additional Figures

Figure C1: Share of Subcategories of NACE 47 in the CZSO and MC Data

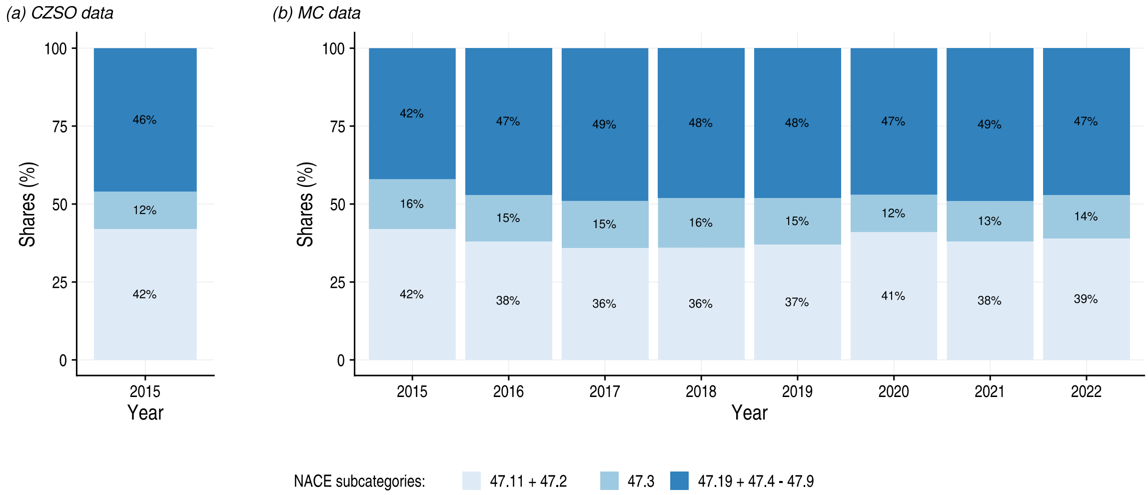
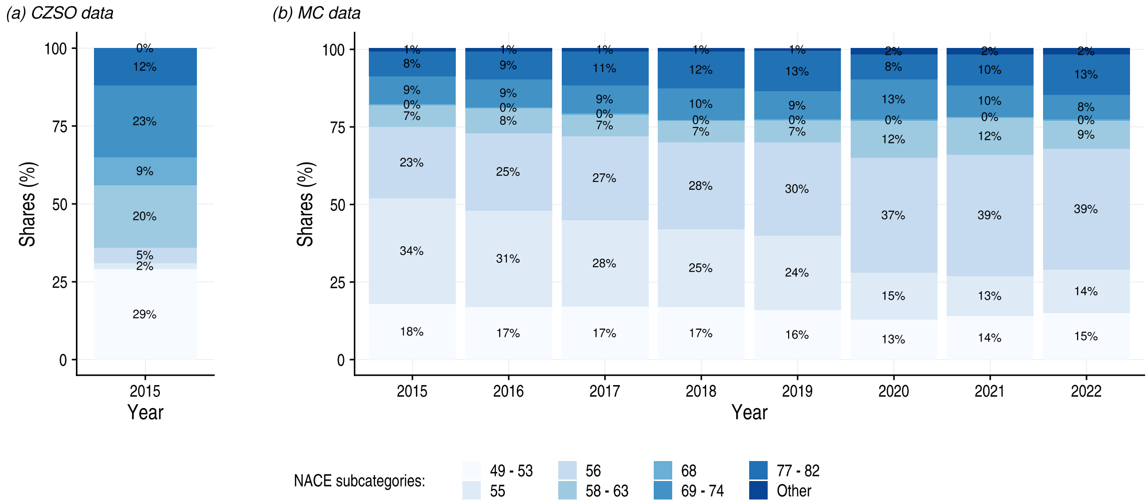


Figure C2: Share of Subcategories of NACE 49–82 (Services) in the CZSO and MC Data



CNB Working Paper Series (since 2022)

WP 14/2023	Tomáš Adam Jan Bělka Martin Hlůže Jakub Matějů Hana Prause Jiří Schwarz	*Ace in hand: The value of card data in the game of nowcasting*
WP 13/2023	Michal Andrle Jan Brůha	*A sparse Kalman filter: A non-recursive approach*
WP 12/2023	Zuzana Gric Jan Janků Simona Malovaná	*What drives sectoral differences in currency derivate usage in a small open economy? Evidence from supervisory data*
WP 11/2023	Dominika Ehrenbergerová Simona Malovaná Caterina Mendicino	*How do climate policies affect holdings of green and brown firms' securities?*
WP 10/2023	Josef Bajzík	*Does shareholder activism have a long-lasting impact on company value? A meta-analysis*
WP 9/2023	Jan Brůha Hana Brůhová Foltýnová	*Long-term impacts of the COVID-19 pandemic on working from home and online shopping: Evidence from a Czech panel survey*
WP 8/2023	František Brázdik Karel Musil Stanislav Tvrz	*Implementing yield curve control measures into the CNB core forecasting model*
WP 7/2023	Alexis Derviz	*Foreign exchange implications of CBDCs and their integration via bridge coins*
WP 6/2023	Simona Malovaná Dominika Ehrenbergerová Zuzana Gric	*What do economists think about the green transition? Exploring the impact of environmental awareness*
WP 5/2023	Milan Szabo	*Cyclical investment behavior of investment funds: Its heterogeneity and drivers*
WP 4/2023	Monika Junicke Jakub Matějů Haroon Mumtaz Angeliki Theophilopoulou	*Distributional effects of monetary policy shocks on wage and hours worked: Evidence from the Czech labor market*
WP 3/2023	Simona Malovaná Jan Janků Martin Hodula	*Macroprudential policy and income inequality: The trade-off between crisis prevention and credit redistribution*
WP 2/2023	Michal Franta	*The Application of multiple-output quantile regression on the US financial cycle*
WP 1/2023	Martin Veselý	*Finding the optimal currency composition of foreign exchange reserves with a quantum computer*

WP 10/2022	Martin Hodula Milan Szabo Josef Bajzík	*Retail fund flows and performance: Insights from supervisory data*
WP 9/2022	Jiří Gregor Jan Janků Martin Melecký	*From central counter to local living: Pass-through of monetary policy to mortgage lending rates in districts*
WP 8/2022	Simona Malovaná Martin Hodula Zuzana Gric Josef Bajzík	*Borrower-based macroprudential measures and credit growth: How biased is the existing literature?*
WP 7/2022	Martin Časta	*How credit improves the exchange rate forecast*
WP 6/2022	Milan Szabo	*Meeting investor outflows in Czech bond and equity funds: Horizontal or vertical?*
WP 5/2022	Róbert Ambriško	*Nowcasting macroeconomic variables using high-frequency fiscal data*
WP 4/2022	Jaromír Baxa Jan Žáček	*Monetary policy and the financial cycle: International evidence*
WP 3/2022	Martin Hodula Milan Szabo Lukáš Pfeifer Martin Melecký	*Cooling the mortgage loan market: The effect of recommended borrower-based limits on new mortgage lending*
WP 2/2022	Martin Veselý	*Application of quantum computers in foreign exchange reserves management*
WP 1/2022	Vojtěch Molnár	*Price level targeting with imperfect rationality: A heuristic approach*

CNB Research and Policy Notes (since 2022)

RPN 2/2023	Eva Hromádková Ivana Kubicová Branislav Saxa	*How does interest rate pass-through change over time? Rolling windows and the role of the credit risk premium in the pricing of Czech loans*
RPN 1/2023	Tomáš Adam Aleš Michl Michal Škoda	*Balancing volatility and returns in the Czech National Bank Bank's foreign exchange portfolio*
RPN 2/2022	Jan Filáček Lucie Kokešová Matějková	*Disclosing dissent in monetary policy committees*
RPN 1/2022	Oxana Babecká Kucharčuková Jan Brůha Petr Král Martin Motl Jaromír Tonner	*Assessment of the nature of the pandemic shock: Implications for monetary policy*

CZECH NATIONAL BANK
Na Příkopě 28
115 03 Praha 1
Czech Republic

ECONOMIC RESEARCH DIVISION
Tel.: +420 224 412 321
Fax: +420 224 412 329
http://www.cnb.cz
e-mail: research@cnb.cz

ISSN 1803-7070