Nowcasting economic activity using transaction payments data

Laura Felber, Simon Beyeler

SNB Working Papers
1/2023

SCHWEIZERISCHE NATIONALBANK
BANQUE NATIONALE SUISSE
BANCA NAZIONALE SVIZZERA
BANCA NAZIUNALA SVIZRA
SWISS NATIONAL BANK

Legal Issues

DISCLAIMER

The views expressed in this paper are those of the author(s) and do not necessarily represent those of the Swiss National Bank. Working Papers describe research in progress. Their aim is to elicit comments and to further debate.

ISSN 1660-7716 (printed version)
ISSN 1660-7724 (online version)

Nowcasting Economic Activity Using Transaction Payments Data

Laura Felber* and Simon Beyeler†‡

April 11, 2023

Abstract

In this paper, we assess the value of high-frequency payments data for nowcasting economic activity. Focusing on Switzerland, we predict real GDP based on an unprecedented 'complete' set of transaction payments data: a combination of real-time gross settlement payment system data as well as debit and credit card data. Following a strongly data-driven machine learning approach, we find payments data to bear an accurate and timely signal about economic activity. When we assess the performance of the models by the initially published GDP numbers (pseudo real-time evaluation), we find a state-dependent value of the data: the payment models slightly outperform the benchmark models in times of crisis but are clearly inferior in 'normal' times. However, when we assess the performance of the models by revised and more final GDP numbers, we find payments data to be unconditionally valuable: the payment models outperform the benchmark models by up to 11% in times of crisis and by up to 12% in 'normal' times. We thus conclude that models based on payments data should become an integral part of policymakers' decision-making.

Keywords: Nowcasting, GDP, machine learning, payments data, COVID-19

JEL classification: C52, C53, C55, E37

*University of Basel and Swiss National Bank (SNB); laura.felber@snb.ch

†Swiss National Bank (SNB); simon.beyeler@snb.ch

‡We thank Elliot Beck, Martin Brown, Ajit Desai, Thibaut Duprey, Alain Galli, Matthias Gubler, Matthias Jüttner, Sébastien Kraenzlin, Sarah Lein, Christoph Meyer, Kurt Schmidheiny, Josine Zanoli and Michael Zimmert for helpful comments and discussions. We would also like to thank the anonymous referee as well as participants of conferences and seminars for their remarks. The views, opinions, findings, and conclusions or recommendations expressed in this paper are strictly those of the authors. They do not necessarily reflect the views of the Swiss National Bank (SNB). The SNB takes no responsibility for any errors or omissions in, or for the correctness of, the information contained in this paper.

1 Introduction

For policymakers, it is essential to have accurate and timely information about the state of the economy. This is particularly true in times of crisis, such as during financial crises or during the recent COVID-19 pandemic, which severely disrupted the global economy. However, 'traditional' macroeconomic indicators, such as the gross domestic product (GDP), are often only published in low frequency (often quarterly), with a significant publication lag and are often subject to significant revisions. Therefore, to assess the current state of the economy, policymakers need to rely on nowcasts. Those, in turn, are often based on high-frequency financial market data, which only partly reflect the state of the real economy, and/or past data that are unable to capture (sharp) economic fluctuations in a timely manner. Hence, there is a need for indicators built with alternative, high-frequency data available in 'real time'. Payments data belong to the set of such alternative data. They reflect the money flow of not only financial but also real activity and thus are closely related to the business cycle. Moreover, they are of high frequency (one observation for each transaction), not subject to revisions and available in almost real-time (in general with a lag of only one business day).

In this paper, we assess the value of an unprecedented 'complete' and granular set of transaction payments data for nowcasting, backcasting and short-term forecasting real Swiss economic activity on the grounds of real GDP.[1] We are the first to work with both transactional payment system and card data, with the aim of improving the prediction of Swiss economic activity. On the one hand, we base our predictions on transaction data from Switzerland's real-time gross settlement (RTGS) system for payments in Swiss francs, the SIC system (SIC stands for 'Swiss Interbank Clearing'), and on the other hand on a large set of transactional debit and credit card data. Whereas debit and credit card payments mostly reflect point of sale (POS) and e-commerce consumption, RTGS payments reflect consumption more broadly as well as gross capital formation and net exports. Moreover, RTGS data comprise financial market signals that are potentially helpful for capturing sharp economic fluctuations. Hence, the two data sources complement each other and together have the potential to predict GDP fluctuations

[1] In what follows, we refer to predictions of the most recent month's GDP number as 'nowcasts', to predictions of past months' (not yet published) GDP numbers as 'backcasts' and to predictions of future GDP numbers as (short-term) 'forecasts'.

in non-turbulent times as well as during crises of various natures – including financial as well as consumption and health crises such as the COVID-19 crisis.

From a methodological point of view, we introduce a strongly data-driven approach and apply machine learning techniques to cope with the high-dimensional and rather 'short' payments data that covers roughly 20 years (RTGS data) respectively 10 years (card data) only. Moreover, we show how to preprocess the highly granular and transactional payments data so that they can realize their full potential as model input. To this end, we apply data cleansing and filter misleading signals of economic growth. In particular, we remove breaks due to institutional changes and control for shifts in payment behavior. We then build a large set of monthly payment streams along various dimensions available in the data. To make predictions, we set up payment models with the payment streams and their lags as predictors. As the target variable, we use the so-called 'MFIGDP', which is a sophisticated interpolation of quarterly Swiss GDP.[2] Finally, we train and test the payment models in a large hyperparameter space based on a variety of linear and non-linear methods (standard and regularized regressions, tree-based methods and artificial neural networks). We further add an optional pre-fitting dimensionality reduction step. Training and testing are performed in an expanding window cross-validation procedure suitable for time series data. To assess the performance of the payment models and, more generally, the value of payments data for predicting GDP, we refer to two benchmarks. On the one hand, we set up direct benchmark models that contain only the target variable's lags as explanatory variables and pursue training and testing with the procedure outlined above.[3] On the other hand, we draw on a practical and more challenging benchmark, the predictions of MFIGDP. MFIGDP thus has a dual role within this paper: it not only provides us with a monthly target (GDP interpolation) but also with a challenging benchmark (GDP predictions).

Our results show that payments data bear an accurate and timely signal about economic activity and are valuable for the prediction of GDP. When we assess the performance of the

[2]MFIGDP matches past quarterly GDP numbers accurately (see Section 2.1). Using a monthly interpolation of official quarterly GDP numbers, i.e., conducting a monthly rather than quarterly prediction exercise, allows us to triple the number of observations, which is reasonable given the 'short' payments data.

[3]Note that the set of models specified by the hyperparameter space also nests the simple autoregressive distributed lag model that is often drawn on as basic benchmark in prediction exercises.

payment as well as both benchmark models by the first vintages of the target (pseudo real-time evaluation), we find a horizon- and state-dependent value of payments data for predicting real economic activity. We find that in the latest COVID-19 crisis, the payment models outperform the direct and practical benchmark for all horizons except $h > 1$. The performance varies with the prediction horizon and is up to 19% better than the performance of the direct benchmark, and up to 5% better than the performance of the practical benchmark. In 'normal' times, the payment models only outperform the backcasts of the direct benchmark and perform worse than the practical benchmark for any horizon (by up to -49%). In contrast, when we assess the performance of payment models as well as the practical benchmark by the latest instead of the real-time vintages of the target, we find a state-independent – but still horizon-dependent – value of the payments data: the root-mean-squared errors (RMSEs) of the back- and nowcasts of the payment models are up to 11% lower than the RMSEs of the practical benchmark in the latest COVID-19 crisis but also up to 12% lower in 'normal' times. This highlights the advantage of payments data not being revised and thus not underlying any measuring errors regarding the current state of the economy.

Overall, we conclude that our adjustments to the raw data, the specified payment models and the data-driven and machine learning based training and testing procedure successfully extract the relevant signal from the high-dimensional and rather 'short' data. We thus infer payments data to be valuable for predicting economic activity. Hence, we suggest that payments data become an integral part of policymakers' information set for decision-making. This can be accomplished by including payments data in existing prediction models or by including alternative and more 'traditional' data sources in the strongly data-driven procedure suggested in this paper.

Indeed, alternative high-frequency data, including payments data, have been increasingly examined for the prediction of economic statistics and have been shown to be valuable for many countries. Examples of recent work cover Norway, the U.S., Italy and Canada. Aastveit, Fastbo, Granziera, Paulsen, and Torstensen (2020) apply a set of MIDAS regressions to nowcast Norwegian quarterly household consumption using weekly and monthly credit and debit card payment streams. Aladangady et al. (2021) develop daily and geographically granular estimates of spending in the U.S. based on electronic payment transaction data by applying multiple stages of filtering, aggregation and transformation. The resulting indices can be understood as a timelier, more granular and higher frequent version of the official U.S. Census Bureau's Monthly Retail Trade Survey. Antolin-Diaz, Drechsel, and Petrella (2021) develop a Bayesian dynamic factor model to compute daily estimates of U.S. GDP growth based on many features and incorporating high-frequency data, including credit card spending. Aprigliano, Ardizzi, and Monteforte (2019) use monthly Italian payment system data and apply a mixed-frequency dynamic factor model to target both quarterly GDP and its main components. Brave, Fogarty, Karger, Aaronson, and Krane (2021) set up a mixed-frequency dynamic factor model and create a weekly index of retail trade for the U.S. that is representative of the U.S. Census Bureau's Monthly Retail Trade Survey and, among others, based on credit and debit card transactions. Chapman and Desai (2020) apply machine learning methods to Canadian payment system data and demonstrate that such data in combination with machine learning techniques are valuable for nowcasting a variety of macroeconomic indicators. Delle Monache, Emiliozzi, and Nobili (2021) compute a weekly index of Italian economic activity based on the first principal component of daily and monthly alternative data, including payment system data.[4] Last, Galbraith and Tkacz (2018) use a large set of Canadian payments data comprising debit card and cheque transactions to nowcast GDP and retail sales and assess the marginal contribution of the data over time.

Two studies concerned with nowcasting Swiss economic activity using payments data shall be highlighted. First, Eckert, Kronenberg, Mikosch, and Neuwirth (2020) set up a mixed-frequency dynamic factor model that incorporates a broad set of time series, consisting of both alternative data (including aggregated numbers of debit and credit card transaction volumes) and 'traditional' macroeconomic and financial data. The extracted factor can be interpreted as a weekly growth rate of Swiss GDP. Second, Wegmüller, Glocker, and Guggia (2023) create a novel index of weekly economic activity in Switzerland that is based on the methodology

[4] Delle Monache et al. (2021) borrow from the methodology of Lewis, Mertens, Stock, and Trivedi (2022) that build a weekly index of economic activity for the U.S. incorporating alternative data that is of purely weekly frequency but not based on any payments data. Similar to Lewis et al. (2022), Delle Monache et al. (2021) scale the index to GDP growth.

of comparable indices in other countries. The index incorporates a variety of high-frequency data, such as the net tonne-kilometers on the Swiss Federal Railways, electricity consumption, cash withdrawals and aggregations of debit and credit card transactions.

Given the current state of research, the value and potential of payments data for predicting economic activity has not yet been fully assessed and exploited. Compared to other studies using payments data to nowcast economic activity, and in particular Swiss economic activity, we work with a transactional, more granular and more 'complete' set of payments data comprising two complementing data sources – RTGS payment system and card data – and focus strongly on their potential. We strive to maximally exploit the informational content in the payments data and contribute to the literature by increasing the methodological knowledge of how to process transaction payments data and, based on the prepared data, how to set up a data-driven and machine learning based training and testing procedure that is capable of capturing economic activity.

Section 2 describes MFIGDP and its dual role as a target (monthly GDP interpolation) and challenging practical benchmark (monthly GDP predictions) as well as the payments data and data preparation in more detail. Section 3 gives a formal specification of the prediction problem and elaborates on the training and testing procedure. Section 4 summarizes our findings. Finally, section 5 concludes.

2 Data

2.1 Monthly GDP: interpolation and predictions

Our target variable is a monthly interpolation of Swiss GDP. Swiss GDP numbers are published annually by the Swiss Federal Statistical Office (FSO) and quarterly by the State Secretariat for Economic Affairs (SECO). To obtain monthly frequency, we use a monthly indicator that is published SNB internally by the Economic Affairs unit. The so-called 'MFIGDP' is related to the SNB's business cycle index (BCI, see Galli (2018)). Both MFIGDP and BCI are based on a mixed-frequency dynamic factor model estimated on a large set of monthly and quarterly –

not payments related – indicators such as labor market, consumption, investment and financial market indicators. MFIGDP matches past quarterly calendar and seasonally adjusted real GDP numbers accurately (due to time aggregation rules). In addition, MFIGDP's observations, which are released before official quarterly GDP is published, provide relatively strong predictions. Figure 1 shows the interpolation and predictions of MFIGDP and illustrates the dual role of the indicator.

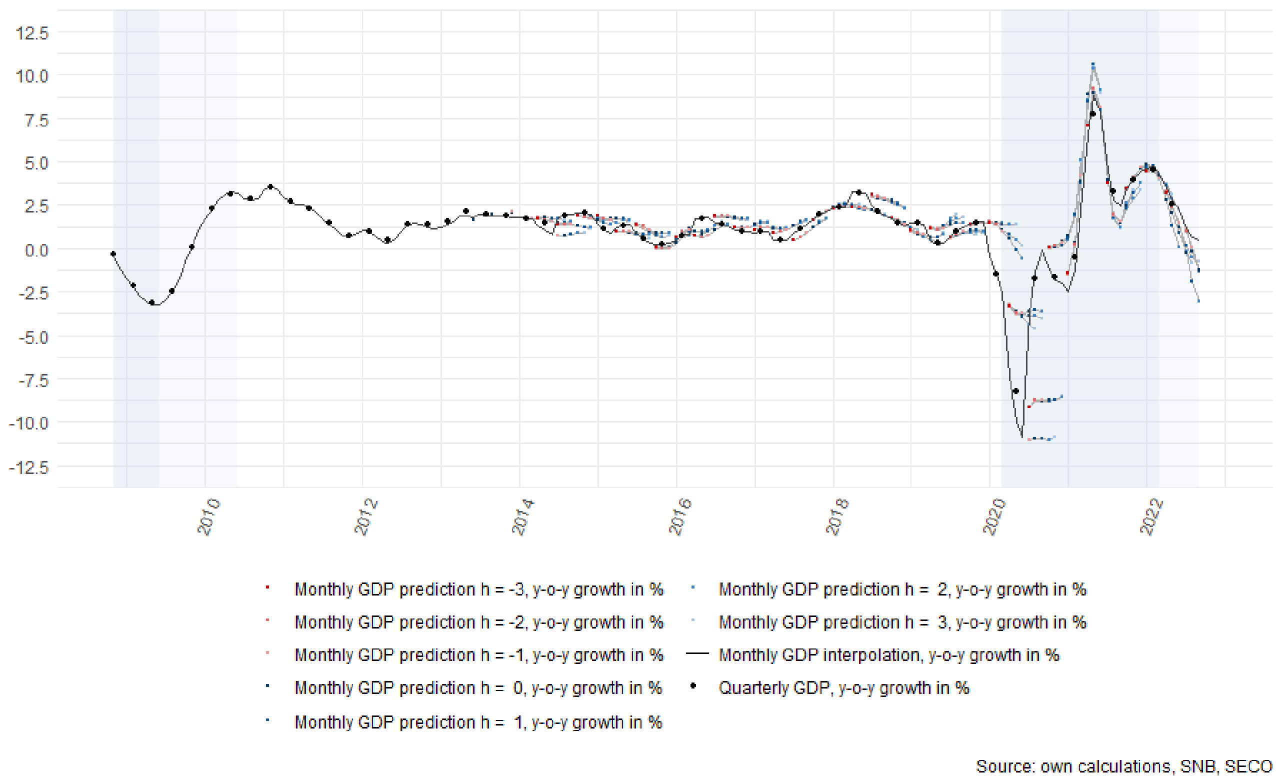

Figure 1: The monthly GDP interpolation (black solid line) accurately matches quarterly calendar and seasonally adjusted real GDP numbers (black dots). Vintages are available from July 2013 onward and provide relatively strong predictions (colored dots; different prediction horizons of one prediction vintage connected by gray lines) in 'normal' times and times of crisis (shaded areas).

By using a monthly interpolation of official GDP numbers, we conduct a monthly rather than a quarterly prediction exercise. On the one hand, a higher-frequency target is more interesting in times of turmoil with sharp and fast economic fluctuations. On the other hand, tripling the number of observations in our prediction exercise is reasonable given the rather 'short' payments data.

In our analysis, we use data covering the time period from January 2005 to September

2022 and transform the data to year-on-year (y-o-y) growth rates. MFIGDP is available from 1990 onward, but we limit our observations to the observation periods of the payments data (see Sections 2.2 and 2.3). MFIGDP's vintages are available from July 2013. MFIGDP's interpolation, i.e., the first vintages of MFIGDP after the official quarterly GDP has been published, serve as target variable and the target's 12 lags as predictors within our prediction exercise (only feasible lags are included in our models, see Subsection 3.2.1). Since payments data date back longer than MFIGDP's vintages, we use the July 2013 vintage for all observations prior to July 2013. MFIGDP's predictions, i.e., the vintages before the publication of the official quarterly GDP serve as a practical and challenging benchmark: we compare the performance of our models' back-, now- and forecasts from July 2013 onward to the performance of MFIGDP's predictions (see Section 4).

2.2 Transactional RTGS payment system data (SIC)

The SIC system is Switzerland's central RTGS payment system and a cooperation between the SNB and the financial center. The SNB acts as a system manager, while the system is operated on behalf of SNB by SIX Interbank Clearing Ltd (SIC Ltd). The SIC system provides unified settlement by settling both interbank payments and retail payments in Swiss francs. Interbank payments are conducted on account of financial institutions themselves, while retail payments are payments banks conduct on behalf and on account of their customers. Interbank payments comprise payments between financial institutions and payments initiated by third-party system operators that are, for example, securities settlement systems and participate in the SIC system by directly effecting debits and credits to the accounts of SIC participants.[5]

The SIC data are transactional and available with a lag of one business day. The data are available from January 2005 onward and are anonymized, such that no conclusions about the payer and payees of retail payments can be drawn. SIC transactions come with a large number of attributes, including a transaction identifier, the date and time of submission and settlement, the amount, the SIC participants involved and the transactions' ISO20022-use

[5]For more information about the SIC system, refer to Felber and Reinke (2022).

cases (message type and payment type). The message type of payments allows the distinction of retail and interbank payments. In the context of interbank payments, the payment type allows, for example, the identification of compensation payments made in the case of a completed transaction (e.g., money market, foreign exchange or securities transaction) in contrast to general payments between two financial institutions (FI-to-FI payments). Moreover, the payment type allows to distinguish between the different third-party operators and the corresponding settlement payments.

We include SIC payment transactions from the first day of January 2005 up to the last day of September 2022 in our analysis and remove transactions that reflect transfers from and to SNB sight deposit accounts as well as transactions from and to one large retail bank participating in the SIC system.[6] Based on the submission date of the transactions, we aggregate the transaction data to a total of 20 monthly payment streams (value and volume) and their 12 lags (resulting in 260 variables): the aggregation is carried out for the total transactions, for interbank as well as retail payments (identified by the message types) and along the different payment types (see Table 1).[7]

The monthly payment streams are transformed into y-o-y growth rates, which has the advantage of removing monthly seasonal patterns (weekly and higher-frequency seasonal patterns are already smoothed out by the aggregation to monthly payment streams). After balancing the dataset (dropping missing values introduced by lagging), we are left with a sample period from January 2007 onward. Appendix A contains figures of the different SIC payment streams together with the target variable (monthly GDP interpolation).

[6] The deposits held by SIC participants in their sight deposit accounts at the SNB are used as the means of payment in the SIC system. At the beginning of a settlement day, SIC participants' sight deposits are transferred from their SNB sight deposit accounts to their settlement accounts in the SIC system. At the end of the settlement day, the deposits are transferred back. Transactions reflecting sight deposit transfers from the SNB sight deposit accounts to the SIC settlement accounts are removed since these transfers are a technical necessity and unrelated to economic activity. In addition, transactions from and to PostFinance Ltd, an important participant in the Swiss payments system, are removed. In 2017, the bank began to gradually integrate bilateral interbank transactions (that had been settled 'inhouse' by the bank (on-us payments)) into the SIC system. The integration was completed in the first half of 2021. Due to the removal of these transactions, we lose a substantial share of retail and overall transactions but the removal allows us to remove the structural break – and the potentially misleading signal of economic growth.

[7] Over the sample period (i.e., 2005 onward), certain message types have been replaced. We thus link old message types to their current counterparts. Moreover, we do not consider payment types that have been discarded or introduced at some point in the period covered.

SIC payment streams	
Total	Total SIC payments (interbank and retail)
Interbank payments	Payments conducted on account of financial institutions themselves, comprising payments between two financial institutions and payments initiated by third-party system operators
Retail payments	Payments conducted on behalf and on account of financial institutions' customers
FI-to-FI payments	General payments between two financial institutions
Compensation payments	Payments made in the case of a completed transaction (e.g., money market, foreign exchange or securities transaction)
SECOM settlement	Payments effected by the corresponding third-party system operator to settle transactions of the Swiss securities settlement system (SECOM) operated by SIC SIS Ltd
Eurex settlement	Payments effected by the corresponding third-party system operator to settle transactions in the Eurex exchange that offers mainly trading in European-based derivatives
Repo settlement	Payments effected by the corresponding third-party system operator to settle repo trades
Bancomat settlement	Payments effected by the corresponding third-party system operator to settle ATM balances between financial institutions
EFT/POS settlement	Payments effected by the corresponding third-party operator to settle point of sale (POS) card transaction balances between financial institutions

Table 1: From the transactional SIC data, we build 20 monthly payment streams (value and volume) and their lags, resulting in 260 variables serving as predictors in the payment models.

2.3 Transactional card data (CARD)

The debit and credit card data comprise both transactions processed by Switzerland's biggest acquirer Worldline Switzerland Ltd (Worldline) (formerly SIX Payment Services) and transactions conducted by cards issued by PostFinance Ltd (PostFinance). Worldline transactions alone account for nearly two-thirds of all card transactions conducted in Switzerland (Kraenzlin, Meyer, and Nellen (2020)). Together with the transactions of PostFinance cards, almost full 'market coverage' is achieved – at least for the years for which both data sources are available (see Figure 2 for indicative representation of 'market coverage').

As with the SIC data, the card data are transactional and available with a lag of one business day. The Worldline data are available from January 2011, and the PostFinance data are available from April 2018. The data are anonymized such that neither individual merchants, card holders nor any other actor involved in the transaction can be identified. Each card transaction includes a relatively large number of attributes, including a transaction identifier, the transaction date and time, the amount, the origin of the card used for the transaction (domestic vs. foreign) and information about the merchant. The latter include, for example, the merchant's location (postal code or canton) and the sector to which it belongs (indicated by the two-digit NOGA code).[8]

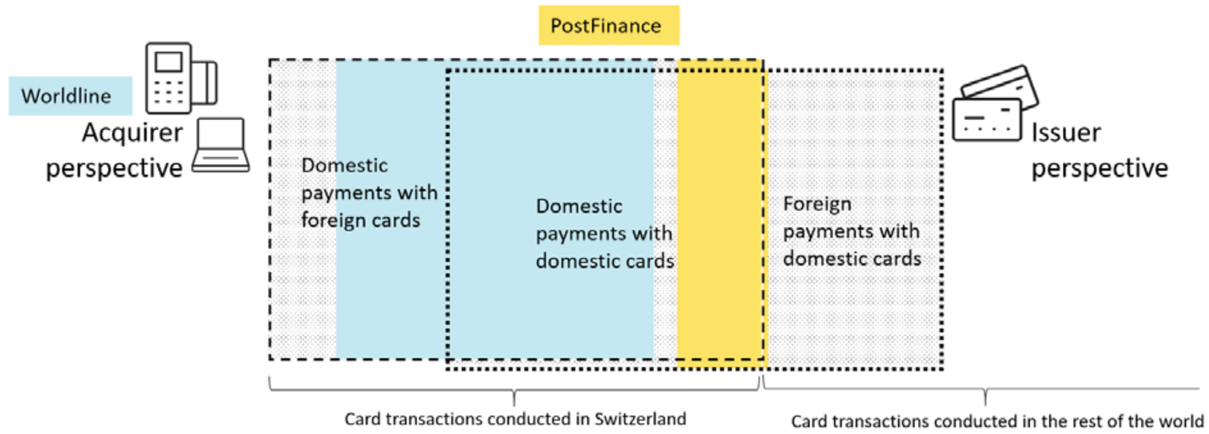

Figure 2: Indicative 'market coverage' of Worldline and PostFinance card data: Worldline and PostFinance card data together lead to almost full 'market coverage' of domestic card payments.

We include Worldline transactions from the first day of January 2011 up to the last day of September 2022 in our analysis. PostFinance data are incorporated from the first day of April 2018 until the last day of September 2022. Based on the transaction date, we aggregated transaction data to a total of 48 monthly payment streams (value and volume) and their 12 lags (resulting in 624 variables): the aggregation is carried out for the total of transactions, transactions conducted by foreign cards and along the different sectors indicated by the two-digit NOGA code (see Table 2).[9]

As with the other data, the card data are transformed into y-o-y growth rates, which handles seasonality. To combine PostFinance data with Worldline data without creating a break in the resulting series and to resolve the problem of shifts in market shares of acquires in our merchant-centric data, we base the y-o-y growth rate calculations for a specific month t only on transactions from merchants for which we observe transactions in both months t and $t-12$. This proceeding is similar to the 'constant-merchant'-approach proposed by Aladangady et

[8]NOGA is the General Classification of Economic Activities and the Swiss version of the European classification of economic activity NACE.

[9]Note that NOGA payment streams with missing observations as well as payment streams showing sudden level shifts are removed.

al. (2021).[10] After balancing the dataset, we are left with data from January 2013 onward. Appendix B contains figures of the different CARD payment streams together with the target variable (monthly GDP interpolation).

CARD payment streams	
Total	Total CARD payments
Foreign card payments	Payments conducted by non-domestic cards
NOGA 41 payments	Payments in the NOGA 41 sector: construction of buildings
NOGA 45 payments	Payments in the NOGA 45 sector: trade and repair of motor vehicles/motorcycles
NOGA 47 payments	Payments in the NOGA 47 sector: retail trade, except of motor vehicles/motorcycles
NOGA 49 payments	Payments in the NOGA 49 sector: land transport, transport via pipelines
NOGA 51 payments	Payments in the NOGA 51 sector: air transport
NOGA 55 payments	Payments in the NOGA 55 sector: accommodation
NOGA 56 payments	Payments in the NOGA 56 sector: food and beverage service activities
NOGA 59 payments	Payments in the NOGA 59 sector: motion picture, video/television program production, sound recording, music publishing activities
NOGA 61 payments	Payments in the NOGA 61 sector: telecommunications sector
NOGA 65 payments	Payments in the NOGA 65 sector: insurance, reinsurance and pension funding, except compulsory social security
NOGA 68 payments	Payments in the NOGA 68 sector: real estate activities
NOGA 69 payments	Payments in the NOGA 69 sector: legal and accounting activities
NOGA 75 payments	Payments in the NOGA 75 sector: veterinary activities
NOGA 79 payments	Payments in the NOGA 79 sector: travel agency, tour operator reservation service and related activities
NOGA 85 payments	Payments in the NOGA 85 sector: education
NOGA 86 payments	Payments in the NOGA 86 sector: human health activities
NOGA 90 payments	Payments in the NOGA 90 sector: creative, arts, entertainment activities
NOGA 91 payments	Payments in the NOGA 91 sector: libraries, archives, museums, other cultural activities
NOGA 93 payments	Payments in the NOGA 93 sector: sports activities, amusement, recreation activities
NOGA 96 payments	Payments in the NOGA 96 sector: other personal service activities
NOGA 99 payments	Payments in the NOGA 99 sector: activities of extraterritorial organizations/bodies

Table 2: From the transactional card data, we build 48 monthly payment streams (value and volume) and their lags, resulting in 624 variables serving as predictors in the payment models.

[10] Aladangady et al. (2021) develop daily and geographically granular estimates of spending at retailers and restaurants in the U.S. based on electronic payment transaction data. In a multistage process, they filter, aggregate and transform the input data into daily spending indices. One important filtering stage is keeping 'constant merchants' only and by doing so correcting their card data for shifts in market shares of the payment processor providing them with the data.

2.4 Cash-card ratio (CCR)

Shifts in payment behavior might lead to a misleading signal of growth within our payment streams. To control for the ongoing shift away from cash to card payments, and in particular the strong acceleration of that development during the beginning of the COVID-19 pandemic, we construct a cash-card ratio (CCR) and include its monthly y-o-y growth rate in models comprising card data as explanatory variables (see Subsection 3.2.1).

The CCR is constructed using data on cash withdrawals and card payments stemming from SNB's surveys on cashless payment transactions and ATMs.[11] The monthly survey data are available from 2005 onward but we limit its observation period to the observation period of the CARD data (see Section 2.3). The survey data, among others, comprise the value and volume of card payments and cash withdrawals at ATMs and POS, broken down into credit card, debit card and e-money transactions and the location of transactions (domestic vs. foreign) as well as the card origin (domestic vs. foreign). The series on cash withdrawal shows a substantial break in series in May 2017, going back to an ATM system migration. Before May 2017, cash withdrawals with cards that belong to the same bank as the ATM (cash withdrawals at ATMs of the card-issuing bank) had not been included in the survey.

We build the CCR using the volumes of debit and credit card payments and cash withdrawals at ATMs and points of sale conducted by domestic and foreign cards in Switzerland. We rely on volumes instead of values because volumes are less sensitive to cash withdrawals made for non-transactional purposes (cash as store of value) that are not related to economic activity per se. We add the cash withdrawals and – to correct for the break in series in May 2017 – subtract cash withdrawals at card-issuing banks and divide the resulting aggregate by the total of card payments.[12] As with the other data, we transform the CCR into y-o-y growth rates.[13] Since the survey data are published with a lag of approximately two months,

[11]The data are accessible on SNB's data portal (last accessed on January 27, 2023) in the table section 'Financial market'. Information about the surveys on cashless payments and ATMs can be found at the SNB's website on survey documents (last accessed on January 27, 2023).

[12]We thank our colleagues from the statistics department for the provision of the series of cash withdrawals at card-issuing banks.

[13]The resulting CCR is a reasonable proxy but of course still has some limitations. First, SNB's surveys on cashless payment transactions and ATMs are partial sample surveys that cover a large part but not the entire 'market'. Moreover, by subtracting cash withdrawals at card-issuing banks, we correct for the mentioned

we extrapolate the data by simply rolling forward its previous values to match the observation period of the CARD data. Figure 3 shows the y-o-y growth rates of the CCR.

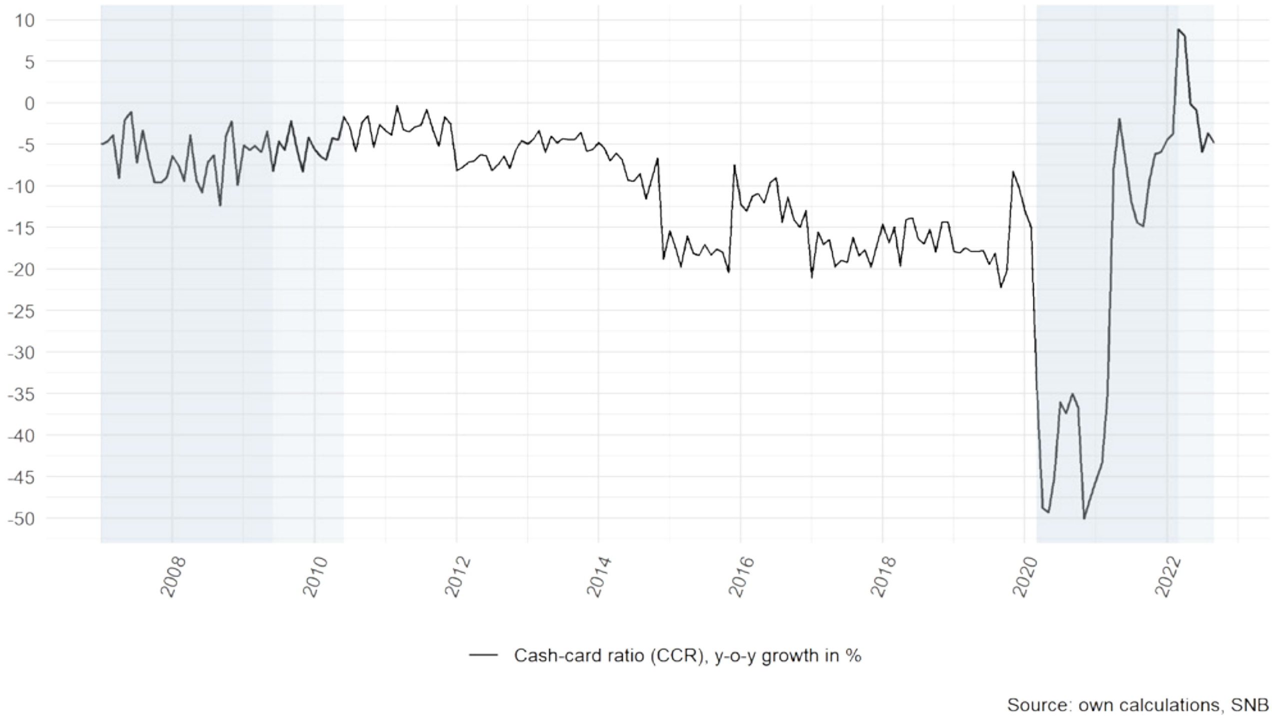

Figure 3: The cash-card ratio (CCR) is constructed from SNB's survey on cashless payment transactions and ATMs. The CCR did not change remarkably during the great financial crisis (shaded area on the left) but had been declining on an ongoing basis ever since and showed a massive drop during the COVID-19 pandemic (shaded area on the right). To control for shifts in payment behavior, and in particular shifts from cash to card payments, the CCR is included in models comprising card data as explanatory variables.

3 Empirical approach

3.1 Formal specification of the nowcasting problem

The methodological representation of the nowcasting problem is straightforward: For a given month $t \in [1, T]$, the target variable (also: dependent variable), i.e., the monthly GDP interpolation, is denoted by y_{t+h} and nowcasted if $h = 0$, forecasted if $h > 0$ or backcasted if $h < 0$. The set of predictors (also features, independent variables or explanatory variables)

break in series but lose a substantial share of cash withdrawals. The calculated ratio is thus based on even fewer (but still many) observations. Second, the data comprise cash withdrawals made for non-transactional purposes (cash as store of value) not reflecting economic activity per se. Third, and as stressed by Dalhaus and Welte (2021), cash withdrawals underestimate cash use since they do not reflect that bank notes and coins are being used in several transactions.

includes the lags of the target variable as well as the different payment streams and their lags (see Tables 1 and 2). A single predictor is denoted by $x_{v,l,t}$ with $v \in [1, V]$ denoting the target if $v = 1$ and a payment stream if $v \neq 1$ and $l \leq L$ being the corresponding lag. The prediction problem is $\boldsymbol{y_h} = f(\boldsymbol{X}) + \boldsymbol{\epsilon}$ and the resulting predictions are given by $\boldsymbol{\hat{y}_h} = f(\boldsymbol{X})$ – with $\boldsymbol{X}$ being the matrix of predictors, $\boldsymbol{y_h} = [y_{1+h}, ..., y_{t+h}, ..., y_{T+h}]^\intercal$ the vector of the target variable, $\boldsymbol{\hat{y}} = [\hat{y_{1+h}}, ..., \hat{y_{t+h}}, ..., \hat{y_{T+h}}]^\intercal$ the corresponding vector of predictions of the target variable at horizon h and $\boldsymbol{\epsilon} = [\epsilon_1, ..., \epsilon_t, ..., \epsilon_T]^\intercal$ the vector of the estimation errors.

From a methodological point of view, using a monthly interpolation of official quarterly GDP numbers, i.e., conducting a monthly rather than quarterly prediction exercise, is reasonable given the rather 'short' payments data – the RTGS data covers roughly 20 and the card data roughly 10 years – since it allows us to triple the number of observations. Nevertheless, the higher-frequency target does not resolve all limiting factors stemming from the low frequency of official GDP. To avoid target leakage, when setting up the predictor matrix $\boldsymbol{X}$, one has to bear in mind that only information that would have been available to the forecaster at the time of prediction is used. As we only want to include the realizations of monthly GDP interpolation for which the quarterly GDP has been published, not all lags of the target variable are 'feasible' predictors and therefore have to be removed. The reason for this is straightforward: we do not want to include information that comes in through the predictions of the dynamic factor model used to construct the monthly GDP and of the predictions of the monthly GDP that serve us as a practical benchmark.

Official quarterly Swiss GDP numbers are published with a delay of approximately two months after a quarter is completed. For simplicity, we assume them to be available on the last day of the second month after each completed quarter.[14] Respecting this publication lag, the lags of the monthly GDP interpolation are 'feasible' with a lag of 2 up to 4 months only, depending on the month and as recorded in Table 3. The payment streams (SIC and CARD), in contrast, are available in almost real time, and no publication lags have to be considered.[15]

[14]In the last 15 years, quarterly Swiss GDP has been published between four days before and six days after the end of the second month of the completed quarters.

[15]Note that the CCR is based on survey data and would also be available with a lag of approximately two months only. For simplicity, however, we assume it to be instantly available. As described in Section 2.4, we extrapolate the CCR with its previous values to match the observation period of the SIC and card data.

Month t	Explanatory variable/s	Lags 0	1	2	3	4	5	6	7	8	9	10	11	12
January / April / July / October	Monthly GDP interpolation	x	x	x	x	o	o	o	o	o	o	o	o	o
	Payment streams	o	o	o	o	o	o	o	o	o	o	o	o	o
February / May / August / November	Monthly GDP interpolation	x	x	o	o	o	o	o	o	o	o	o	o	o
	Payment streams	o	o	o	o	o	o	o	o	o	o	o	o	o
March / June / September / December	Monthly GDP interpolation	x	x	x	o	o	o	o	o	o	o	o	o	o
	Payment streams	o	o	o	o	o	o	o	o	o	o	o	o	o

Table 3: To avoid target leakage, when setting up our monthly-frequency models, we have to respect the implicit publication lag of the official quarterly GDP: the lags of the monthly GDP interpolation are 'feasible' with a lag of 2 up to 4 months only, depending on the month t that is predicted and as recorded in this table.

Moreover, the maximum horizon that needs to be backcasted also depends on month t, as specified in Table 4. If t equals, for example, February, backcasting January is meaningful, while backcasting December and November is not since the December and November values of the target are already known and included as explanatory variables (lags) in the prediction model.

Month t	Horizons -3	-2	-1	0	1	2	3
January / April / July / October	o	o	o	o	o	o	o
February / May / August / November	x	x	o	o	o	o	o
March / June / September / December	x	o	o	o	o	o	o

Table 4: Respecting the implicit publication lag of official quarterly GDP numbers, the maximum horizon that can be backcasted depends on the month t. If t equals for example February, backcasting January is meaningful, while backcasting December and November is not since the December and November values of the target are already known.

With the monthly GDP as the target, it follows that all lags $l \in [0, L]$ if $v \neq 1$ and the lags $l \in [2, L]$, $l \in [3, L]$ and $l \in [4, L]$ if $v = 1$ and depending on the month $t \in [1, T]$, are feasible. In our training and testing procedure we either use the full $\boldsymbol{X}$ or apply a dimensionality reduction step (see Subsection 3.2.3). Accordingly, $\boldsymbol{X}$ would be a $T \times V(L + 1) - 4$-matrix if $t \in \{\text{January, April, July, October}\}$ and if all variables $v \in [1, V]$ and feasible lags would be included, i.e., no further dimensionality reduction was conducted:

$$
\boldsymbol{X} = \begin{bmatrix}
x_{1,4,1} & \dots & x_{1,l,1} & \dots & x_{1,L,1} & x_{2,0,1} & \dots & x_{2,l,1} & \dots & x_{2,L,1} & \dots & x_{v,l,1} & \dots & x_{V,L,1} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
x_{1,4,t} & \dots & x_{1,l,t} & \ddots & x_{1,L,t} & x_{2,0,t} & \dots & x_{2,l,t} & \dots & x_{2,L,t} & \dots & x_{v,l,t} & \dots & x_{V,L,t} \\
\vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
x_{1,4,T} & \dots & x_{1,l,T} & \dots & x_{1,L,T} & x_{2,0,T} & \dots & x_{2,l,T} & \dots & x_{2,L,T} & \dots & x_{v,l,T} & \dots & x_{V,L,T}
\end{bmatrix}
$$

3.2 Training and testing procedure

Figure 4 schematically summarizes our strongly data-driven training and testing procedure, which comprises three main elements. **Element 1** is the specification of the payment and direct benchmark models. We are evaluating the potential of two different payments data sources available for different time periods. The SIC payment streams date back to 2007 and hence contain the great financial crisis, while the CARD data streams are only available from 2013 onward. Therefore, we set up a 'long' payment model using only the SIC payment streams and a 'short' payment model that comprises both the SIC and CARD payment streams. As a direct benchmark, we accordingly train and test two models – a 'long' and a 'short' model – that contain only the target variable's lags as explanatory variables but are trained following the same procedure as our payment models. In addition to the direct benchmark models, we compare the performance of our models to the predictions of the monthly GDP (MFIGDP predictions, see Section 2.1). The monthly GDP predictions serve as a practical and rather challenging benchmark. Moreover, as a robustness check, we apply our procedure for $h = 0$ (nowcasting) to quarterly numbers using official quarterly Swiss GDP and quarterly aggregated payment streams.

Element 2 is the specification of the training universe. First, we specify a set of statistical methods that are used to fit our payment and direct benchmark models. Concretely, we focus on four established machine learning methods: elastic net, random forest, gradient boosting and (shallow) artificial neural networks. With these four methods, we cover a broad range of models, including standard linear and regularized regressions as well as tree-based models and artificial neural networks that allow for non-linear relations. Then, for each of these methods, we specify a grid of hyperparameters. Since we cannot tell ex ante which are the best hyperparameters, we 'let the data select' what hyperparameter combinations work best by cross validation (element 3). Moreover, although the chosen machine learning methods should in principle be able to handle high dimensionality (except for the standard linear regressions as part of elastic net), we not only fit our payment models with all explanatory variables but additionally introduce a pre-fitting dimensionality reduction step. To this end, we fit all of our our models with all explanatory variables but also models with a reduced number determined by applying a ridge regression that precedes the model fitting. We treat the choice of whether a dimensionality reduction step is applied, the number of variables to 'keep' and the corresponding penalty parameter of the ridge regression as additional hyperparameters.

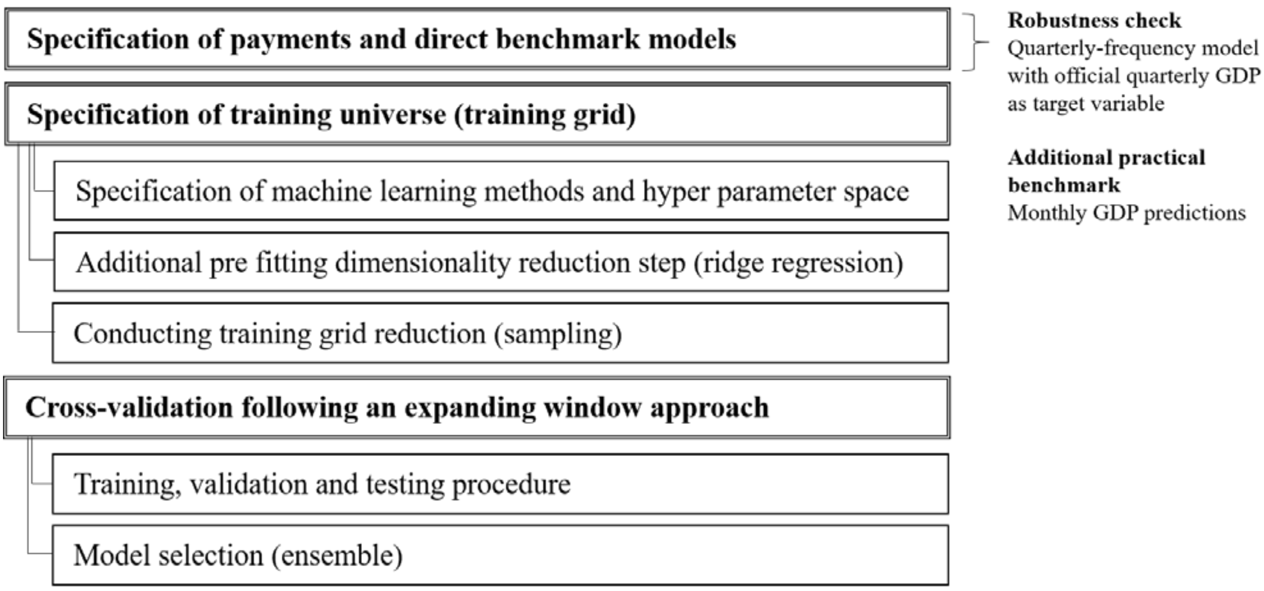

Figure 4: Our strongly data-driven training and testing procedure comprises three main elements: the specification of the models, the specification of the training universe and the actual training and testing following an expanding window cross-validation procedure suited for time series data.

The intersection of all the different hyperparameter ranges leads to over 10 000 distinct models (hyperparameter combinations) that – when matched with all the point in times that shall be predicted and the horizons of the predictions (h) – result in a huge training grid that is no longer very practical given our infrastructure setup. We thus perform a reduction of the grid size by random sampling.

Finally, **element 3** comprises the training of the (still large) training grid for the specified payment and benchmark models, the selection of the best performing models (hyperparameter combinations) and the testing of the selected model ensemble. We do so by following an expanding window cross-validation approach applicable to time series data. In what follows, we further elaborate on the mentioned sub elements of the training and testing procedure.

3.2.1 Specification of payment and direct benchmark models

Using the representation $\boldsymbol{y_h} = f(\boldsymbol{X}) + \boldsymbol{\epsilon}$ from Section 3.1, we specify the following payment models for prediction:

$\boldsymbol{y_h}$: GDP; $\boldsymbol{X}$: GDP, SIC, CARD, CCR (payment model CARD-SIC)

$\boldsymbol{y_h}$: GDP; $\boldsymbol{X}$: GDP, SIC (payment model SIC)

As explained in Section 2.4, to control for shifts in payment behavior and in particular shifts from cash to card payments, models with CARD payment streams as explanatory variables also comprise a non-lagged series reflecting the evolution of the monthly y-o-y growth of cash withdrawals in comparison to card transactions (cash-card-ratio, CCR). Moreover, the number of observations the models are trained on equals the number of observations of the 'shortest' data source incorporated in a model. Consequently, the model with SIC and CARD payment streams as explanatory variables (payment model CARD-SIC) is trained on data from January 2013 onward; the model comprising SIC payment streams only (payment model SIC) is trained on data from January 2007 onward.

The direct benchmark model only incorporates the feasible lags of the target variable itself and, to ensure comparability, is trained with both data back to January 2007 and data back to January 2013 ('long' and 'short' benchmark models).

$\boldsymbol{y_h}$: GDP; $\boldsymbol{X}$: GDP ('short' and 'long' direct benchmark models)

Note that, as explained in Section 3.1, GDP as part of $\boldsymbol{X}$ takes three different states, depending on what month t is to be predicted and as a consequence of the publication lag of quarterly Swiss GDP. Note further that when training and testing the models, all variables are standardized.

3.2.2 Specification of machine learning methods and hyperparameter space

We form predictions by employing the following supervised machine learning methods: elastic net, random forest, gradient boosting and artificial neural networks.[16] For each of these methods, we specify a grid of hyperparameters, including, for example, the mixing parameter alpha and the penalty parameter lambda for the elastic net or the number of trees for random forest and gradient boosting.[17] We suggest referring to Hastie, Tibshirani, and Friedman (2009) for an introduction to the four machine learning methods and their hyperparameters.

3.2.3 Specification of additional pre-fitting dimensionality reduction step

For our payment models, we train the specified models (hyperparameter combinations) with all (feasible) explanatory variables but also with a reduced number of variables by applying a pre-fitting dimensionality reduction step. We do not apply a dimensionality reduction step to our direct benchmark models since they only incorporate the target's lags and thus are already relatively low-dimensional. For the models that we choose to run with a dimensionality reduction step, we conduct – embedded in the expanding window cross-validation procedure (see Subsection 3.2.5) – stacked model fitting. In a first step, we fit our target on the

[16]For elastic net, we use the R package glmnet (version 4.1.1) by Friedman, Hastie, and Tibshirani (2010), for random forest the R package ranger (version 0.13.1) by Wright and Ziegler (2017), for gradient boosting the R package gbm (version 2.1.8) by Greenwell, Boehmke, Cunningham, and GBM Developers (2020) and for artificial neural networks keras (version 2.9.0) by Allaire and Chollet (2022).

[17]For elastic net, we specify the hyperparameter grid $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$ for the mixing parameter alpha and $\{0, 0.02, 0.06, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5\}$ for the penalty lambda. For random forest, we set the parameter splitrule to 'variance', the parameter importance to 'impurity', the number of trees to 5000 and train our data along the combinations of the hyperparameter default values of the ranger package and the parameters mtry and min.node.size for which we use the values $\{3, 5, 7\}$. For gradient boosting, we chose the default values of the R package and set for the distribution parameter the values 'gaussian', 'laplace' and 'quantile' (with $alpha = 0.5$ (median)), for the interaction depth the values $\{1, 3\}$, for the parameter n.minobsinnode $\{3, 5\}$, for the shrinkage parameter $\{0.001, 0.01, 0.1\}$, for the parameter bag.fraction $\{0.5, 1\}$ and for the parameter n.trees a value of 5000. Finally, for neural networks, we specify dense two-layered neural networks with the first layer comprising 15 units and the second layer comprising one unit. We choose 'relu' activation and define the learning rate to be 0.001. The loss function (lossFunction) used is 'mse'. We specify networks without and with regularization (with a regularization rate of 0.001) as well as with and without dropout (with a dropout rate of 0.2). We then fit the specified networks with 100 epochs, no batching (by setting batch_size to the number of observations) and model checkpoint callbacking.

explanatory variables by applying a ridge regression and select the variables with the highest absolute betas (first-step model selection). In a second step, we fit the models with the reduced number of variables with the specified machine learning model (post-ridge).[18] As mentioned in the preceding subsection, we treat the choice of whether a dimensionality reduction step is applied, the number of variables to 'keep' and the corresponding penalty parameter of the ridge regression as additional hyperparameters.[19]

3.2.4 Conducting training grid reduction (sampling)

The specified hyperparameter space (see Subsection 3.2.2) leads to a huge training grid that is – given our current infrastructure setup – not practical. Hence, we reduce our training grid: we include all the combinations where no pre-fitting dimensionality reduction step is performed and, in addition, sample from the combinations with pre-fitting dimensionality step randomly 1000 distinct hyperparameter combinations for each of the four machine learning methods. This leaves us with a still large but computationally feasible training grid.[20]

3.2.5 Training, validation and testing procedure

When data are independent and identically distributed (i.i.d.), cross-validation is conducted by randomly separating the data into training, validation and test sets. Time series data are not i.i.d. Hence, cross-validation preserving the temporal order of the observations is reasonable (Bergmeir, Hyndman, and Koo (2018)). With time series data, the training set should only contain observations that occurred prior to the observations of the validation and test sets. This avoids target leakage, i.e., using data for training that are not yet realized and giving the model an 'unfair' advantage. For our exercise, we thus apply an expanding window cross-validation approach suitable for time series data.

[18]We chose ridge over lasso because ridge allows us to explicitly control the number of variables remaining in the model. Note that neither of the two regularized regression approaches allows to infer the relative importance of the explanatory variables for the prediction exercise. If variables are highly correlated, the first-step model selection of ridge and lasso is often unstable.

[19]For the number of variables to 'keep', we specify the following values: $(\{1, 10, 15, 20, 25, 30, 50\})$. For conducting the ridge regression, we again use the R package glmnet (version 4.1.1) by Friedman et al. (2010) and set the elastic net mixing parameter alpha to 0. For the penalty parameter of the ridge regression, we specify the same hyperparameter space as before: $\{0, 0.02, 0.06, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5\}$.

[20]Note that for testing, only the selected best performing models are fitted (see Subsection 3.2.6), resulting in a testing grid of a manageable size.

In the same manner as we respected the feasibility of lags of the target variable as explanatory variables in our models (see Section 3.1), we implement the cross-validation procedure by only fitting observations prior to the validation and test sets for which official GDP numbers have already been published. Analogously, only the prediction performance of validation sets prior to a test set that would have been feasible are evaluated for model selection. Different from the 'standard' k-fold cross-validation approach for i.i.d. data, the expanding window cross-validation approach applied does not validate and test the fitted models in one fixed test set (hold-out sample) but sequentially in one test set after another. The expanding window approach thus does not find one global solution but updates its solution with every additional month. For the final evaluation of model performance (see Section 4), the RMSEs over the sequential test sets in times of crisis as well as in times of non-crisis are calculated. The expanding window approach is described in more detail and schematically illustrated in Appendix C.

The randomized k-fold expanding window approach for time series data as proposed by Chapman and Desai (2022) is in limit equivalent to our approach.[21] The randomized k-fold expanding window approach has proven to work well for nowcasting Canadian economic activity using payments data and is similar to the 'Rep-holdout' method, which performed best in a comparison of various cross-validation procedures carried out by Cerqueira, Torgo, and Mozetic (2020).[22]

3.2.6 Model selection (ensemble)

For testing, the models (hyperparameter combinations) are selected by their performance in the validation steps. We measure the performance in terms of the RMSEs of the pseudo out-of-sample predictions across all (feasible) validation steps preceding a test observation. For each model trained (see Subsection 3.2.1) and each horizon, h, we choose the 10 elastic net models, the 10 tree-based models (random forest and gradient boosting) and the 10 artificial

[21] Different from the approach that we apply, the randomized k-fold expanding window approach tests it in a fixed test set (hold-out sample) and, hence, finds a global solution. For the first test observation of the fixed test set, in limit, the procedure is equivalent to our approach.

[22] Cerqueira et al. (2020) compare the 'Rep-Holdout' procedure to other cross-validation and out-of-sample approaches in empirical experiments. The 'Rep-holdout' procedure performed best in the experiment using real-world data.

neural networks with the lowest average RMSE. We then fit these 30 models in the test step and build the mean prediction. By choosing several models per machine learning method and building prediction ensembles, we aim to achieve more robust out-of-sample predictions.

4 Results

4.1 Performance of payment compared to direct benchmark models

When we assess the performance of our payment models and the direct benchmark models by the first vintages of the monthly GDP interpolation (pseudo real-time evaluation), we find a state-dependent value of payments data: payments data tend to be more valuable in times of crisis than in 'normal' times. This is encouraging, as an accurate assessment of the current state of the economy is especially important during times of crisis. Figures 5 and 6 as well as the tables in Appendix D.1 show the performance of the models visually and numerically.

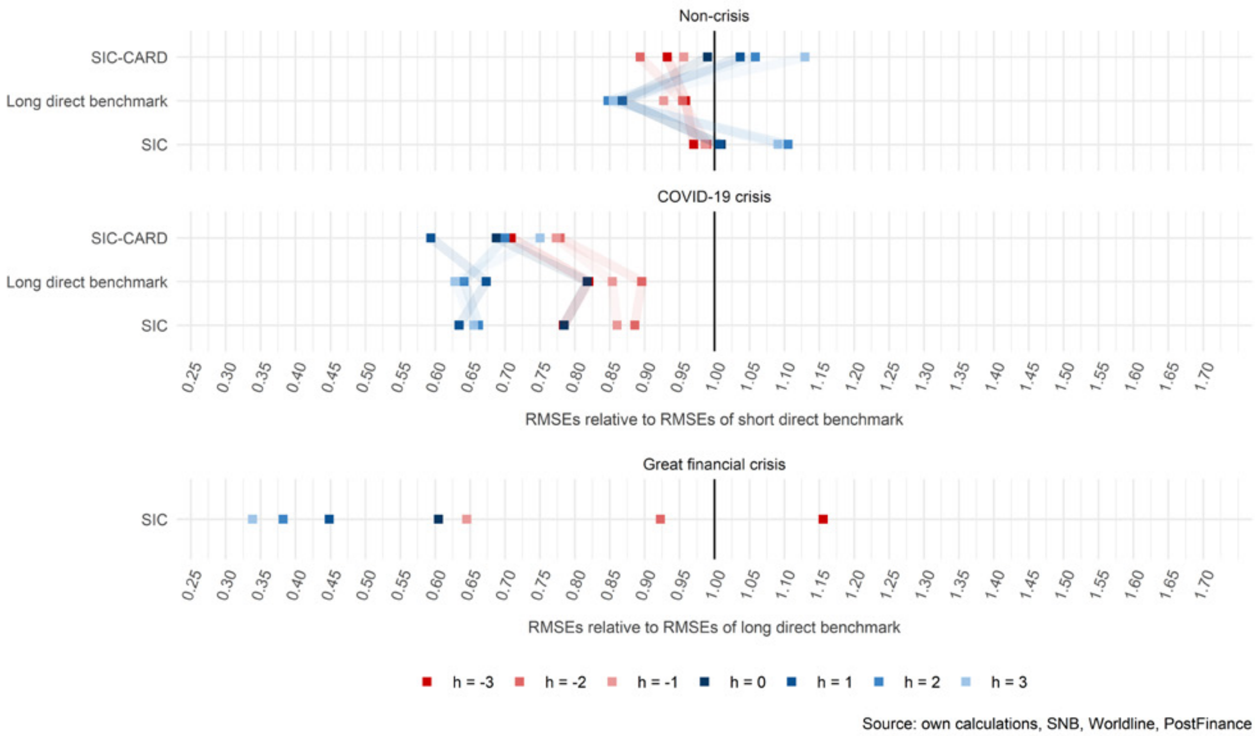

Figure 5: The real-time out-of-sample RMSEs of the payment models relative to the direct benchmark models show the state- and horizon-dependent value of payments data: payments data are valuable in times of crisis but add more noise than value in 'normal' times. Moreover, payments data are more valuable for $h = [\text{-}3, 1]$ than $[2, 3]$.

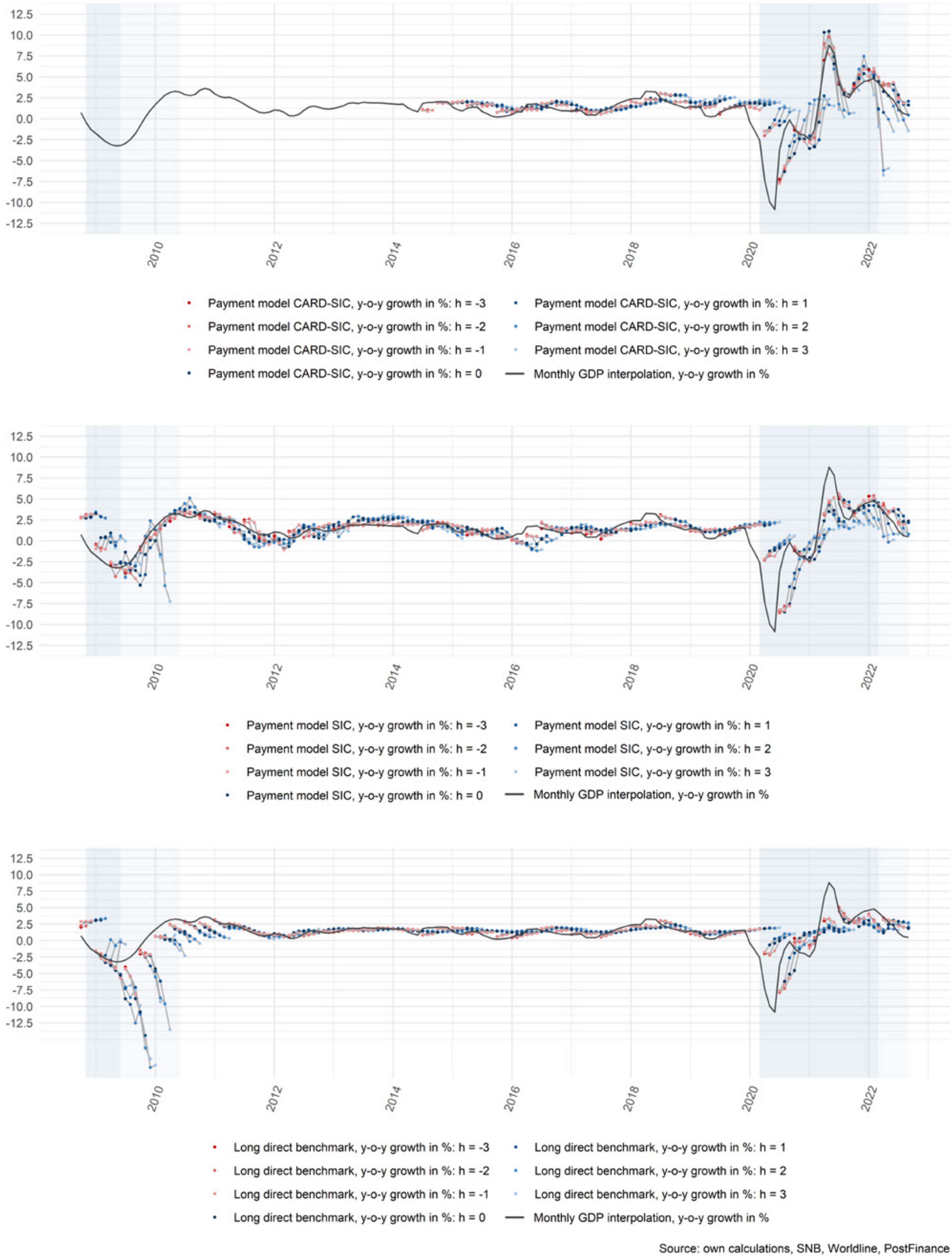

Figure 6: The payment models' out-of-sample back- and nowcasts outperform the direct benchmark models in times of crisis (shaded areas). The payment models perform worse for short-term forecasting. In particular, the forecasts of the payment model CARD-SIC get noisy during transmission to post-pandemic times.

As the figure and tables reveal, the 'long' benchmark is superior to the 'short' benchmark in the overlapping test sets for any state and any horizon. This outperformance can be explained by the fact that the 'long' compared to the 'short' benchmark was trained on a longer data sample, which also includes another crisis. Moreover, the two payment models are substantially superior to the 'short' direct benchmark model during the COVID-19 pandemic (by up to 68%). In addition, payment models, and in particular the payment model CARD-SIC, outperform the 'long' direct benchmark model for most horizons during the COVID-19 pandemic (by up to 19%). Likewise, the payment model SIC clearly outperforms the 'long' benchmark model during the great financial crisis for all horizons except for horizon h = -3 (by up to 195%).[23] On the contrary, in 'normal' non-crisis times, the payment models can only partially keep up with the 'short' benchmark and tend to be inferior to the 'long' direct benchmark (by up to -32%). Put differently, payments data seem to add more noise than value in 'normal' times, which is, for example, also found by Eckert et al. (2020), who conclude that alternative high-frequency data, including debit and credit card transaction volumes, can be valuable for nowcasting GDP in times of turmoil but add noise during more quiet times. Similarly, Chapman and Desai (2022) find payments data to be useful in 'normal' and turbulent times but state a 10% to 20% higher performance reduction (RMSE) in times of crisis compared to 'normal' time.

The value of payments data depends not only on the state (crisis vs. non-crisis), but also on the (type of) crisis and the forecast horizon. Payments data tend to be more valuable for back- and nowcasting than for forecasting. The RMSEs of the payment and direct benchmark models strictly increase with the prediction horizon. This reflects that the prediction exercise becomes more challenging with an increasing forecast horizon. During the COVID-19 pandemic, the RMSEs of both payment models and direct benchmark models strictly increase with the prediction horizon h, but the former at a faster pace than the latter. Consequently, the *relative* RMSEs of the payment models (i.e., the RMSEs relative to the RMSEs of the benchmark models) tend to increase with the horizon h. During the great financial crisis, again, the

[23] The out-of-sample predictions during the great financial crisis are based on relatively few observations only: the minimal training set comprises 18 observations or 1,5 years of data as described in Section 3.2 and Appendix C, respectively. The outstanding performance of payment model SIC compared to the direct benchmark model during the great financial crisis thus illustrates the value of additional explanatory variables, i.e., the value of models comprising more than only past values of the target, which is most pronounced if the data history is short and the data show sharp fluctuations.

RMSEs of the payment model SIC as well as the ‘long’ direct benchmark increase strictly with the horizon h. However, in this crisis, the payment model performs worse than the benchmark for three periods back (h = -3) but clearly outperforms the benchmark for all other horizons and at an increasing rate.

Figures D.1 and D.2 as well as the tables in Appendix D.2 show the results of nowcasting official quarterly Swiss GDP based on the same training and testing procedure with payment streams aggregated to quarterly frequency (robustness check). The quarterly results confirm the state-dependent value of payments data for nowcasting economic activity.

4.2 Performance of payment models compared to practical benchmark

The pseudo real-time assessment of the performance of payment models and the practical benchmark, i.e., the predictions of MFIGDP, confirms the state- and horizon-dependent value of payments data: payments data tend to be more valuable in times of crisis than in ‘normal’ times and improve the performance of back- and nowcasts but not of forecasts (see Figure 7 and the tables in Appendix D.3). Concretely, when we assess the performance of our payment models as well as the practical benchmark by the real-time vintages of the target, we find that the now-, back- and short-term forecasts one period ahead of the payment model CARD-SIC are somewhat superior to the practical benchmark model during the COVID-19 pandemic (by up to 5%). In contrast, the short-term forecasts two and three periods ahead and the predictions of the payment model SIC for most horizons underperform compared to the practical benchmark during the COVID-19 pandemic (by up to -14%). Compared to the ‘long’ direct benchmark (see Section 4.1), the payment models cannot keep up with the practical benchmark in ‘normal’ non-crisis times (being inferior by up to -49%).

However, when we measure the performance of payment models as well as the practical benchmark based on the latest available vintage of the target variable, we nevertheless find payments data to bear a relevant state-independent – but still horizon-dependent – signal. The payment models’ back- and nowcasts of revised and more final monthly GDP numbers outperform those of the challenging practical benchmark in times of crisis but also in times of non-crisis (see Figure 8 and the tables in Appendix D.3). Compared to the real-time RMSEs, the latest vintage RMSEs of the payment model CARD-SIC improve substantially for both states and all horizons (by up to 25% in non-crisis and by up to 15% in crisis times), while the latest vintage RMSEs of the payment model SIC and the practical benchmark generally deteriorate in 'normal' times but slightly improve for some horizons in times of crisis. More precisely, when comparing the latest vintage RMSEs of our payment models to those of the practical benchmark, we find a positive value of payments data for predicting economic activity in times of crisis as well as in 'normal' times: the back- and nowcasts of the payment model CARD-SIC outperform those of the practical benchmark in both states (by up to 12% in non-crisis and by up to 11% in crisis times). Moreover, in times of crisis, the forecast one period ahead of the payment model CARD-SIC also beats the practical benchmark by 11%. However, the payment model CARD-SIC's forecasts two and three periods ahead as well as the payment model SIC's predictions for any horizon perform worse than the practical benchmark.

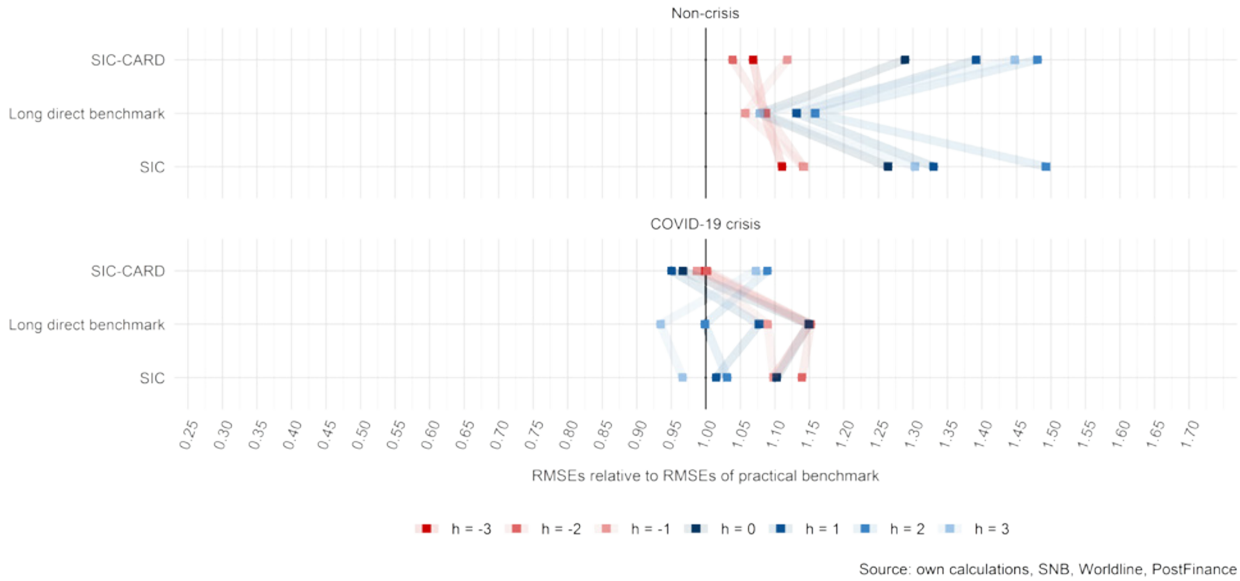

Figure 7: The real-time out-of-sample RMSEs of the payment models relative to the practical benchmark confirm the state- and horizon-dependent value of payments data already found by assessing the payment models' performance relative to the direct benchmark models (see Section 4.1).

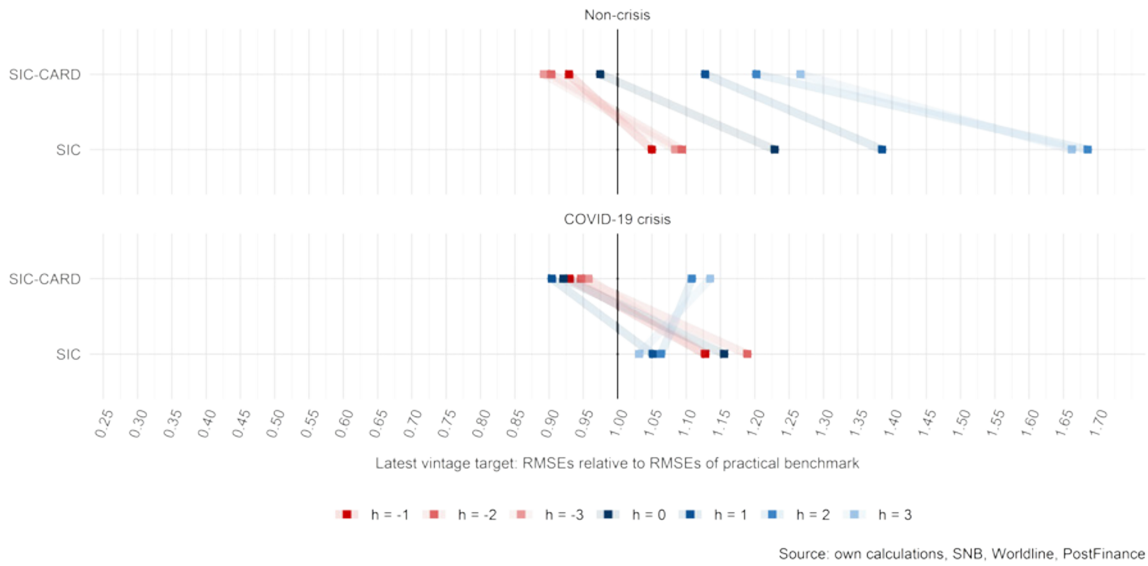

Figure 8: The out-of-sample RMSEs based on the latest vintage of the target (monthly GDP interpolation) relativize the state-dependent value of payments data for predicting GDP: according to that measure, the payment model CARD-SIC's back- and nowcasts ($h = [\text{-}3, 0]$) outperform the practical benchmark's predictions in 'normal' non-crisis and crisis times.

5 Conclusion

In this paper, we assess the value of high-frequency transaction payments data for nowcasting economic activity. We predict GDP using payment streams built from two sources of payments data: RTGS payment system data as well as debit and credit card data. These two data sources complement each other and together reflect significant parts of GDP. To our knowledge, we are the first to work with such a 'complete' and granular set of Swiss payments data. We strive to maximally exploit its informational content by following a strongly data-driven procedure forming prediction based on machine learning methods.

To this end, we apply data cleansing, aggregate the prepared data to a broad set of monthly payment streams, specify payment models comprising the payment streams and their lags as explanatory variables and train and test the payment models with a large variety of linear and non-linear methods. To assess the performance of the payment models and the value of payments data for nowcasting GDP, we draw upon two benchmarks. On the one hand, we compare the performance of the payment models to direct benchmark models, i.e., models that contain only the target variable's lags as explanatory variables but are trained and tested following the same procedure as we apply to the payment models. On the other hand, we compare the payment models' performance to a more practical and more challenging benchmark, the so-called 'MFIGDP', which provides relatively strong GDP predictions at monthly frequency.

Overall, our results show that payments data are valuable for the prediction of economic activity in times of crisis as well as in non-crisis times. When we assess the performance of the payment models and the benchmarks by the first vintages of the target (real-time evaluation), we find a state- and horizon-dependent value of the data. The payment models' back- and nowcasts generally outperform the direct as well as the practical benchmarks in times of turmoil (compared to the direct benchmark by up to 19% during the COVID-19 and by up to 195% during the great financial crisis; compared to the practical benchmark by up to 5% during the COVID-19 crisis). The payment models perform worse for short-term forecasting in times of turmoil and more generally for any prediction horizon in 'normal' non-crisis times (compared to the direct benchmark by up to -32%; compared to the practical benchmark by up to -49%). However, when we assess model performance based on the revised and more final values instead of the real-time vintages of the target, we find a state-independent but still horizon-dependent value of the payments data: the payment models outperform the challenging practical benchmark by up to 11% in the latest COVID-19 crisis but also in non-turbulent times (by up to 12%). This highlights the advantage of payments data over other data sources: they are not revised and have great potential to bear a relevant signal about revised and more final GDP numbers.

A limiting factor that hampers the performance of our models and the found value of payments data to some extent is the low frequency and publication lag of the official quarterly Swiss GDP that translates to our monthly target in combination with the rather short time span of the available payments data.[24] In particular, the card data are rather 'short', and the

[24]Using a monthly measure of GDP instead of official quarterly GDP numbers as target variable allows us to triple the observations that we base our estimations on, but – as discussed in Section 3 – does not resolve all limiting factors stemming from the low frequency of official GDP. Despite its monthly frequency, we only include those months for which the official quarterly GDP has already been published in the training step. As the publication lag of quarterly GDP is approximately 2 months, the implicit publication lag of our monthly GDP measure is 2 to 4 months, depending on the month. To avoid target leakage, this implicit publication lag has to be considered when conducting pseudo out-of-sample nowcasts with past data, i.e.,

payment model CARD-SIC, which is based on both card and RTGS payment system data, does not cover turbulent times before the COVID-19 pandemic. On the positive side, the models might only now be 'trained' for future crises of various nature, and the value of the data for crisis prediction will most likely increase with time. Hence, we conclude that our adjustments to the raw data and our strongly data-driven training and testing procedure successfully extract the relevant signal from the data and find payments data to be valuable for predicting economic activity, in particular for back- and nowcasting in turbulent times.

Our results suggest payments data, i.e., payments data-based models, to become an integral part of policymakers' decision-making. Consequently, central banks and other policy institutions should focus on the combination of payments data with other alternative as well as more 'traditional' data sources and established prediction models, respectively. Either payments data could be included in existing prediction models or other data sources could be included in the strongly data-driven procedure suggested in this paper. The latter might be more promising given the high-dimensionality and rather 'short' history of payments data (as well as of other alternative data). Alternatively, outcomes of established prediction models and more novel approaches could be combined by including the predictions of one model to the other as an additional input variable.

the last 2 to 4 observations must not be used, which considerably limits the nowcasting performance.

References

Aastveit, K. A., Fastbo, T. M., Granziera, E., Paulsen, K. S., & Torstensen, K. N. (2020). Nowcasting Norwegian household consumption with debit card transaction data. *Norges Bank Research*, *17*.

Aladangady, A., Aron-Dine, S., Dunn, W., Feiveson, L., Lengermann, P., & Sahm, C. (2021). From transactions data to economic statistics: Constructing real-time, high-frequency, geographic measures of consumer spending [Book]. In *Big data for twenty-first-century economic statistics.* University of Chicago Press. Retrieved from http://www.nber.org/chapters/c14267

Allaire, J., & Chollet, F. (2022). keras: R interface to 'keras' [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=keras (R package version 2.9.0)

Antolin-Diaz, J., Drechsel, T., & Petrella, I. (2021). Advances in nowcasting economic activity: Secular trends, large shocks and new data. *CEPR Discussion Paper*(15926).

Aprigliano, V., Ardizzi, G., & Monteforte, L. (2019). Using payment system data to forecast economic activity. *International Journal of Central Banking*, *15*(4), 55–80.

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics and Data Analysis*, *120*, 70-83.

Brave, S. A., Fogarty, M., Karger, E., Aaronson, D., & Krane, S. (2021). Tracking U.S. consumers in real time with a new weekly index of retail trade. *Working Paper Federal Reserve Bank of Chicago*, *05*.

Cerqueira, V., Torgo, L., & Mozetic, I. (2020). Evaluating time series forecasting models: an empirical study on performance estimation methods. *Machine Learning*, *109*(1997-2028).

Chapman, J. T., & Desai, A. (2020). Using payments data to nowcast macroeconomic variables during the onset of COVID-19. *Journal of Financial Market Infrastructures*, *9*(1).

Chapman, J. T., & Desai, A. (2022). Macroeconomic predictions using payments data and machine learning. *Staff Working Paper*, *2022-10*.

Dalhaus, T., & Welte, A. (2021). Payments habits during COVID-19: Evidence from high-frequency transaction data. *Staff Working Paper*, *2021-43*.

Delle Monache, D., Emiliozzi, S., & Nobili, A. (2021). Tracking economic growth during the Covid-19: a weekly indicator for italy. *Bank of Italy Note Covid-19, January*.

Eckert, F., Kronenberg, P., Mikosch, H., & Neuwirth, S. (2020). Tracking economic activity with alternative high-frequency data. *KOF Working Papers*(488).

Felber, L., & Reinke, R. (2022). *The Swiss Interbank Clearing (SIC) payment system - report on the SIC system and disclosure report 2021.* Retrieved from https://www.snb.ch/en/mmr/reference/sicsystem_disclosure/source/sicsystem_disclosure.en.pdf

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. Retrieved from https://www.jstatsoft.org/v33/i01/

Galbraith, J. W., & Tkacz, G. (2018). Nowcasting with payments system data. *International Journal of Forecasting*, *34*, 366–376.

Galli, A. (2018). Which indicators matter? Analyzing the swiss business cycle using a large-scale mixed-frequency dynamic factor model. *Journal of Business Cycle Research*, *14*(2), 179–218.

Greenwell, B., Boehmke, B., Cunningham, J., & GBM Developers. (2020). gbm: generalized boosted regression models [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=gbm (R package version 2.1.8)

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction.

Kraenzlin, S., Meyer, C., & Nellen, T. (2020). Covid-19 and regional shifts in Swiss retail payments. *Swiss Journal of Economics and Statistics*, *156*(14).

Lewis, D. J., Mertens, K., Stock, J. H., & Trivedi, M. (2022). Measuring real activity using a weekly economic index. *Journal of Applied Econometrics*, *37*(4), 667-687.

Wegmüller, P., Glocker, C., & Guggia, V. (2023). Weekly economic activity: Measurement and informational content. *International Journal of Forecasting*, *39*(1), 228–243.

Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*(1), 1–17. doi: 10.18637/jss.v077.i01

A Figures of SIC payment streams and monthly GDP interpolation

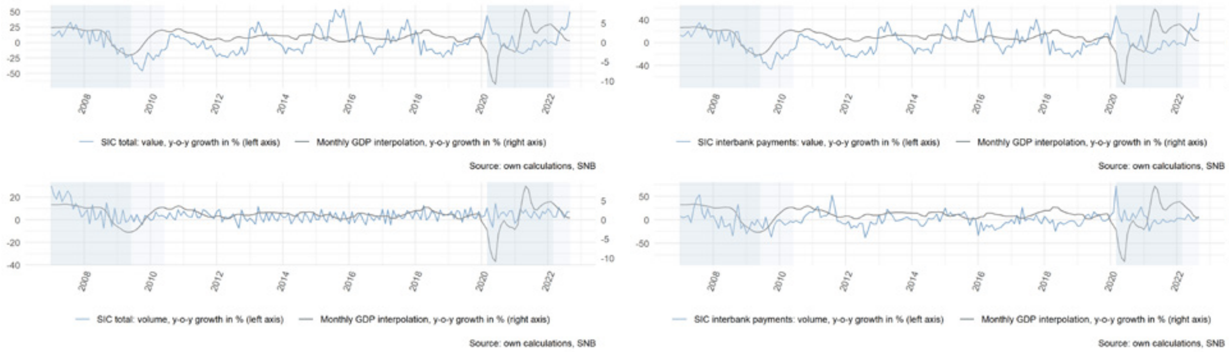

Figure A.1: Total SIC payments (left), interbank payments (right) (value and volume) and monthly GDP interpolation; shaded areas mark times of turmoil (great financial crisis and COVID-19 crisis)

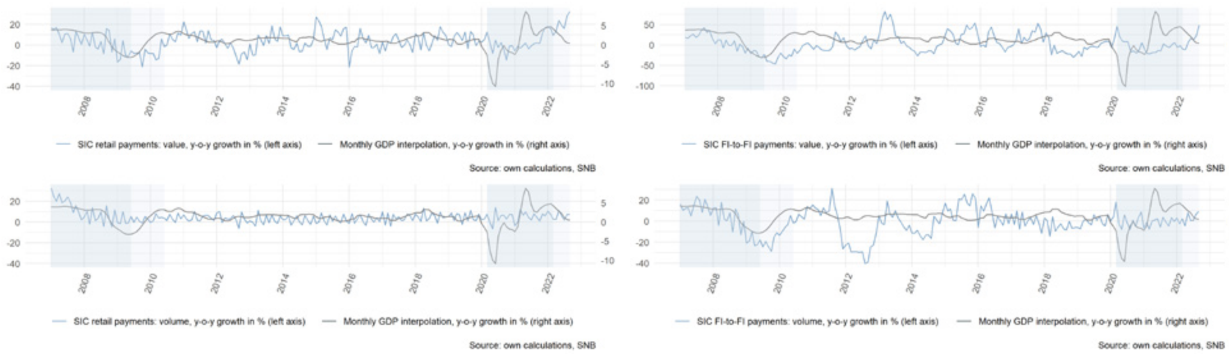

Figure A.2: Retail payments (left), FI-to-FI payments (right) (value and volume) and monthly GDP interpolation; shaded areas mark times of turmoil (great financial crisis and COVID-19 crisis)

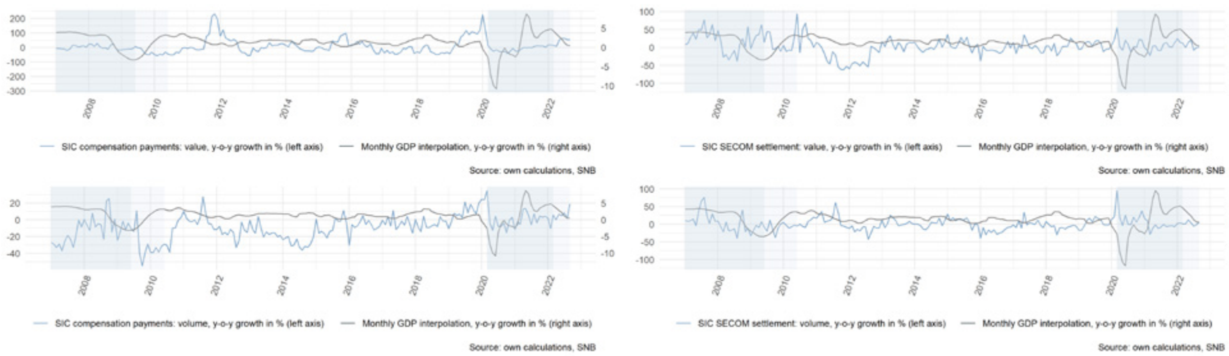

Figure A.3: Compensation payments (left), SECOM settlement payments (right) (value and volume) and monthly GDP interpolation; shaded areas mark times of turmoil (great financial crisis and COVID-19 crisis)

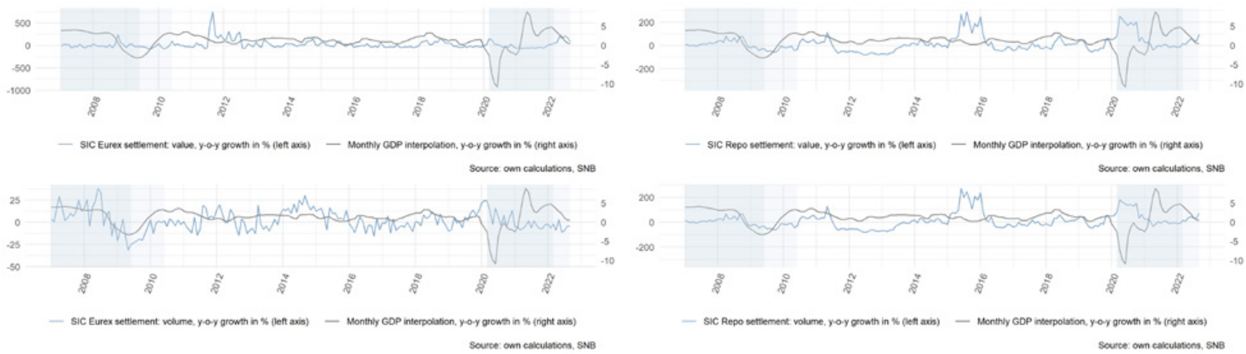

Figure A.4: Eurex settlement payments (left), Repo settlement payments (right) (value and volume) and monthly GDP interpolation; shaded areas mark times of turmoil (great financial crisis and COVID-19 crisis)

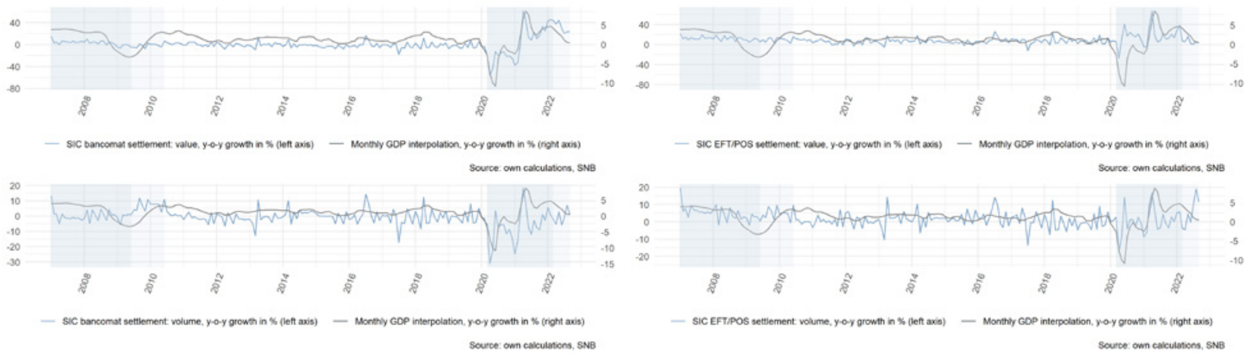

Figure A.5: Bancomat settlement payments (left), EFT/POS settlement payments (right) (value and volume) and monthly GDP interpolation; shaded areas mark times of turmoil (great financial crisis and COVID-19 crisis)

B Figures of CARD payment streams and monthly GDP interpolation

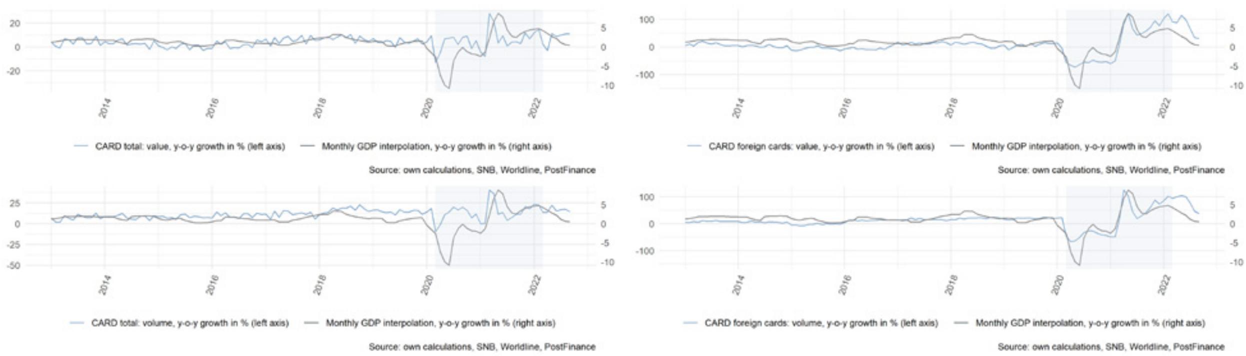

Figure B.1: Total CARD payments (left), foreign card payments (right) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

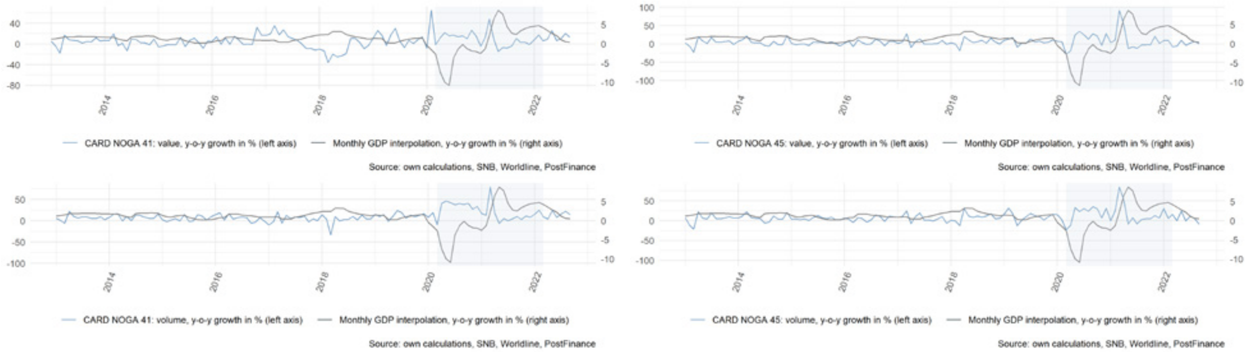

Figure B.2: NOGA 41 payments (construction of buildings; left), NOGA 45 payments (trade and repair of motor vehicles/motorcycles; right) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

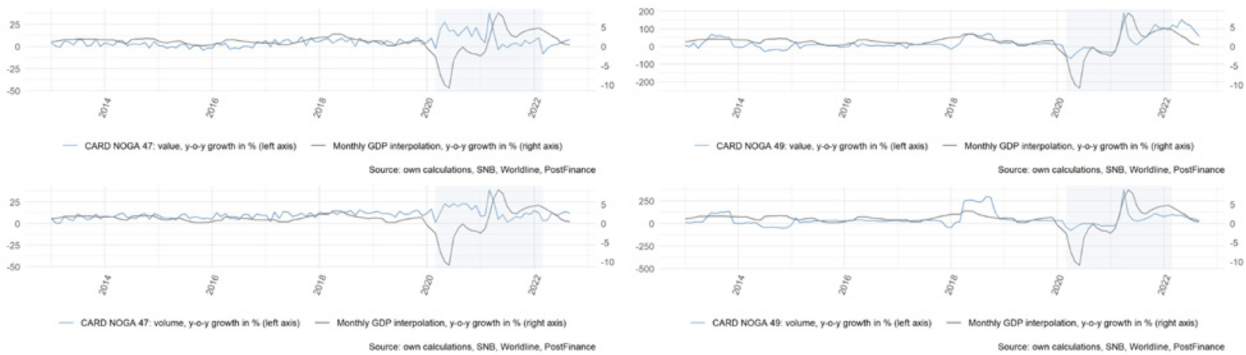

Figure B.3: NOGA 47 payments (retail trade, except of motor vehicles/motorcycles; left), NOGA 49 payments (land transport, transport via pipelines; right) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

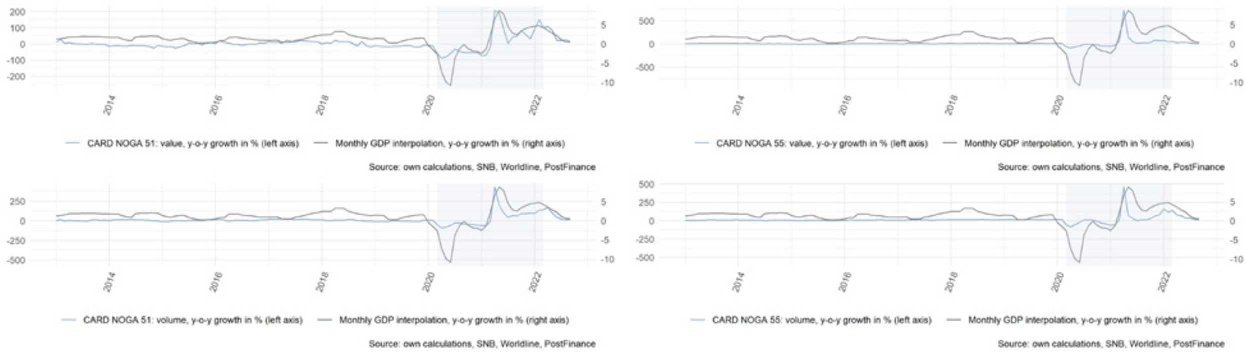

Figure B.4: NOGA 51 payments (air transport; left), NOGA 55 payments (accommodation; right) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

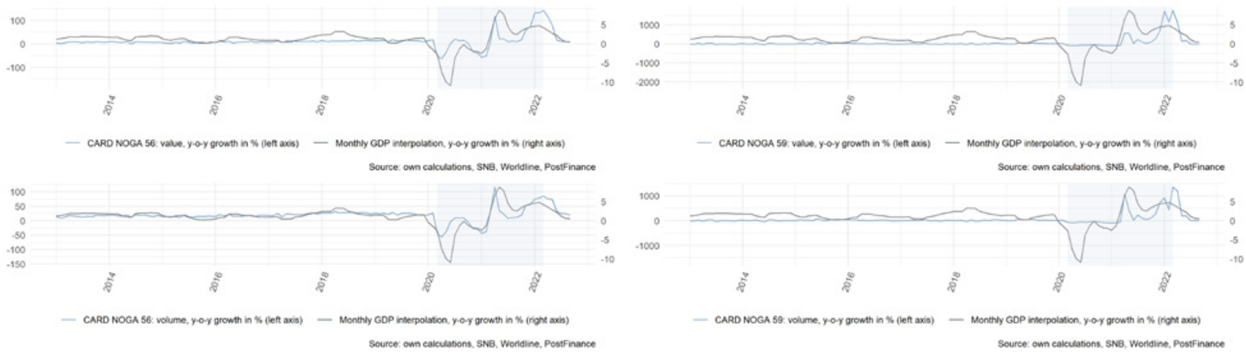

Figure B.5: NOGA 56 payments (food and beverage service activities; left), NOGA 59 payments (motion picture, video/television program production, sound recording, music publishing activities; right) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

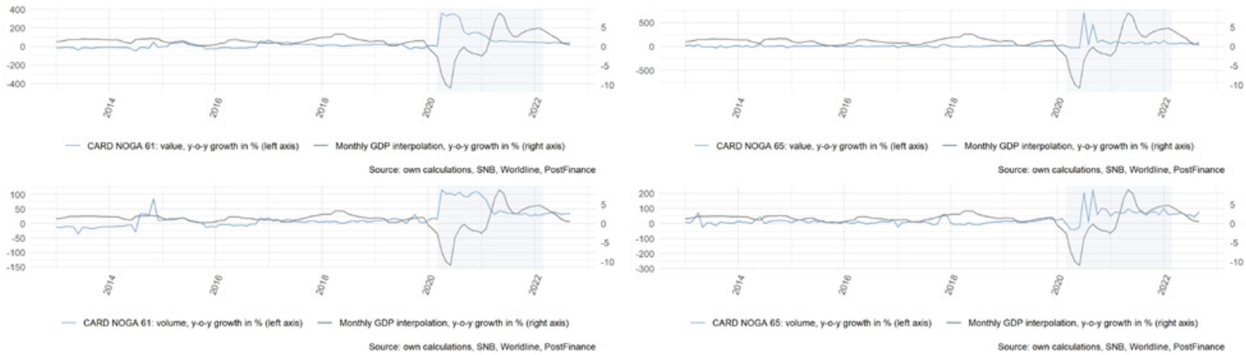

Figure B.6: NOGA 61 payments (telecommunications; left), NOGA 65 payments (insurance, reinsurance and pension funding, except compulsory social security; right) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

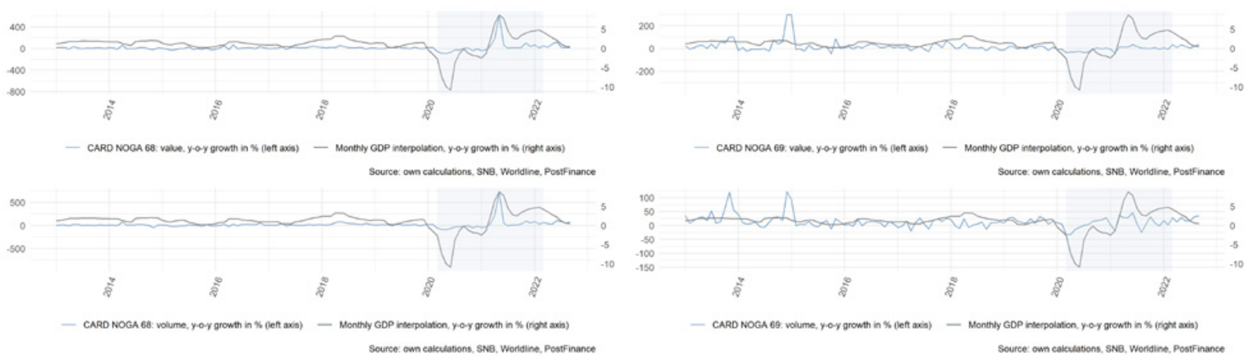

Figure B.7: NOGA 68 payments (real estate activities; left), NOGA 69 payments (legal and accounting activities; right) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

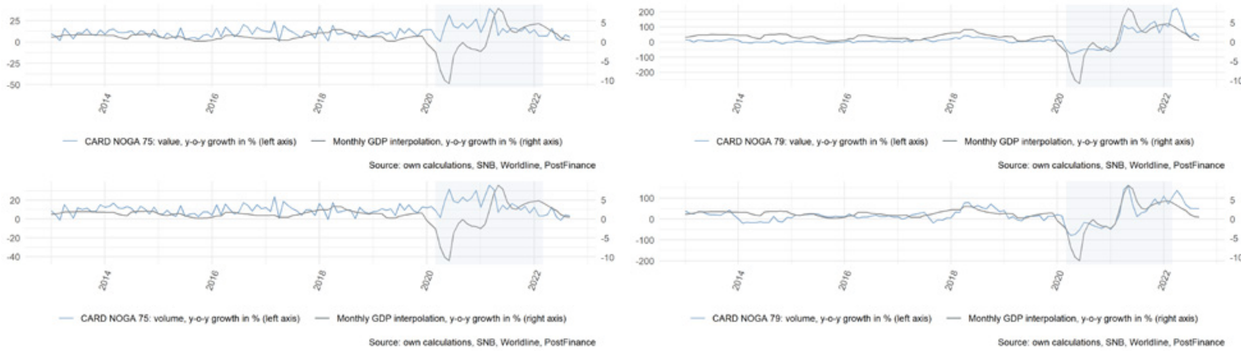

Figure B.8: NOGA 75 payments (veterinary activities; left), NOGA 79 payments (travel agency, tour operator reservation service and related activities; right) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

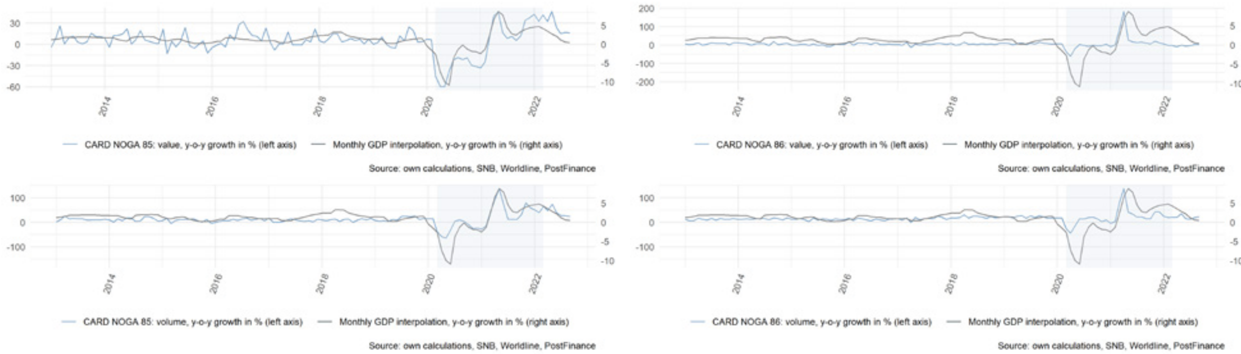

Figure B.9: NOGA 85 payments (education; left), NOGA 86 payments (human health activities; right) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

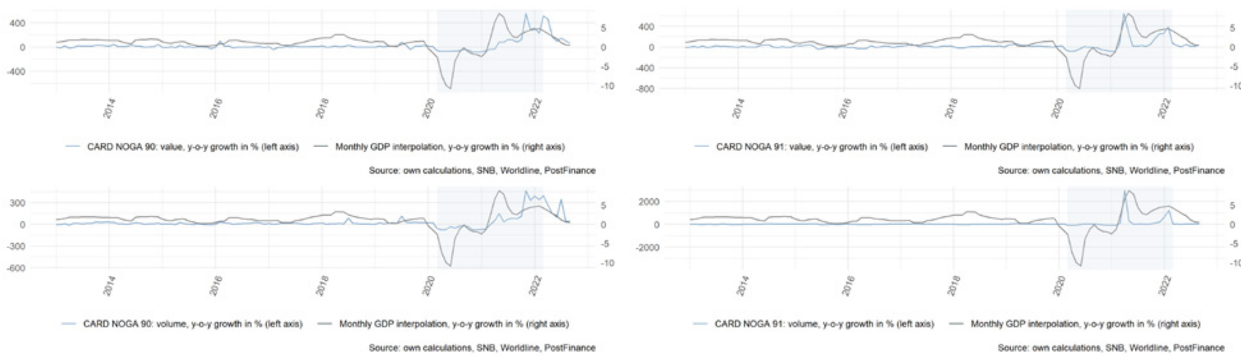

Figure B.10: NOGA 90 payments (creative, arts, entertainment activities; left), NOGA 91 payments (libraries, archives, museums, other cultural activities; right) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

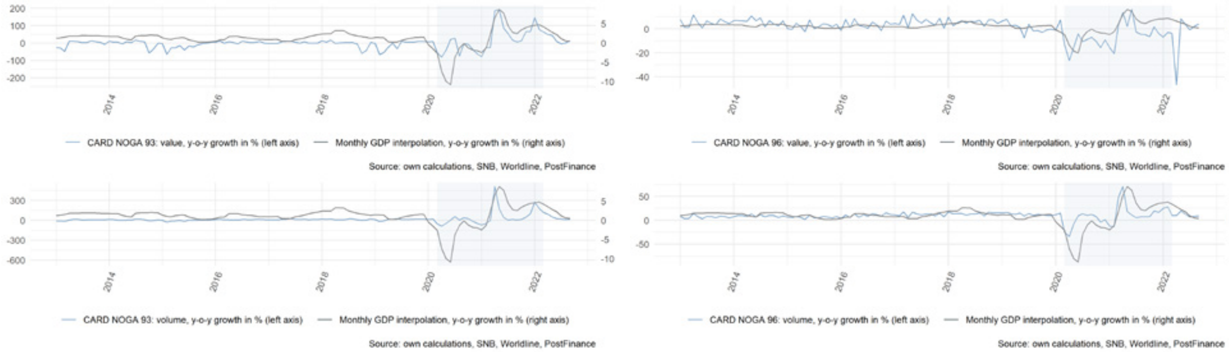

Figure B.11: NOGA 93 payments (sports activities, amusement, recreation activities; left), NOGA 96 payments (other personal service activities; right) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

Figure B.12: NOGA 99 payments (activities of extraterritorial organizations/bodies) (value and volume) and monthly GDP interpolation; shaded area marks the COVID-19 crisis

C Description of the expanding window cross-validation approach for time series data

In the expanding window approach that we follow, the model universe is fitted, validated and tested sequentially (see Figure C.1 for a schematic representation). For training and validation, the first m observations serve as the minimal training set, while the observation following the initial training set serves as the first validation point (observation $m + 1$).[25] The model universe is fitted (trained with the initial training set), and the fitted models are then validated with the initial validation point (pseudo-out-of-sample prediction in the validation set). Then, the training set is extended, and the validation point is moved by one observation (expanding window). Again, the models are fitted with the training data, and the fitted models are

[25]We chose our minimal training set to equal 18 observations ($m = 18$), which correspond to 18 months or 1.5 years. As a robustness check, we train and test our models with the official quarterly GDP as the target and quarterly payment streams as predictors. With quarterly data, our minimal training set also comprises 18 observations, which in that case equal 4.5 years.

validated with the validation point (observation $m+2$). This procedure is repeated until the end of the data is reached. For testing, again an expanding window is followed with observation $m+2$ serving as the first test observation. The best performing model or models, i.e., the model or models with the lowest root-squared errors from the initial validation step are selected and tested with the first test observation (pseudo out-of-sample prediction in the test set). Then, the best performing model or models in the initial and second validation steps are selected based on their RMSEs and tested with the test observation $m+3$. This procedure is repeated until the last observation is reached.[26] To evaluate the performance of our procedure over the whole test set, we calculate the RMSE of the pseudo out-of-sample predictions in the test sets.[27]

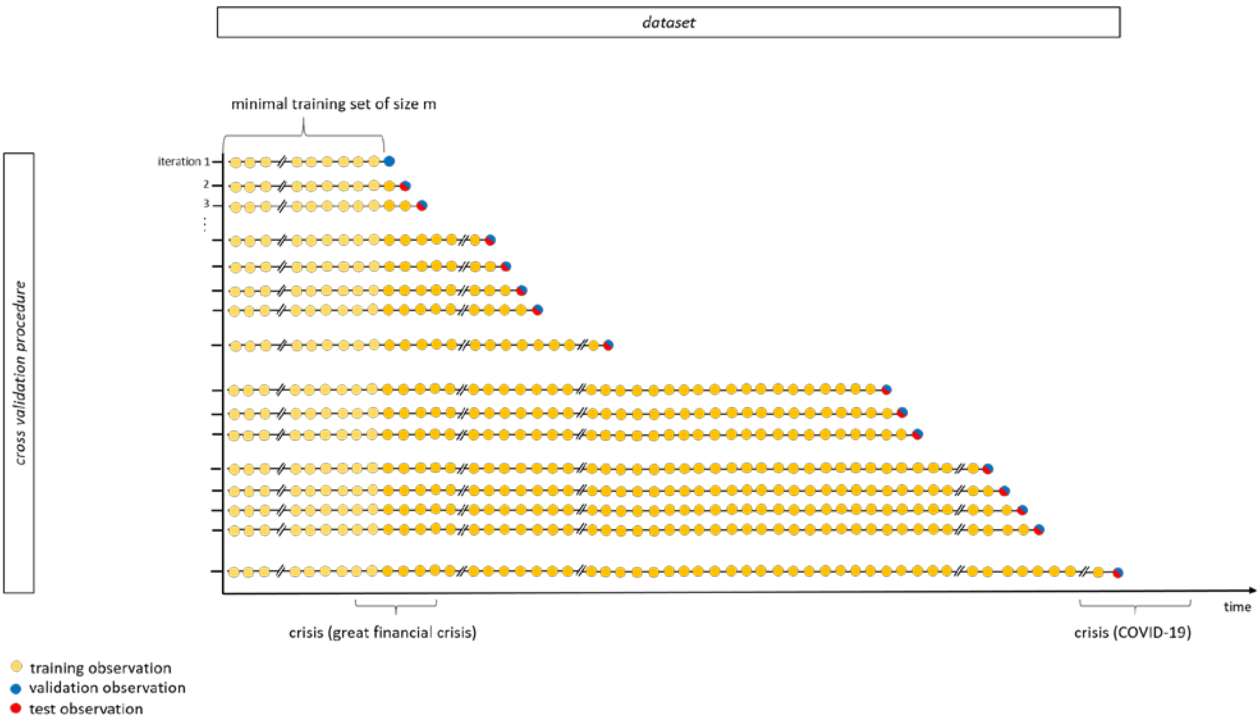

Figure C.1: Schematic representation of the expanding window approach applied in this paper for predictions of horizon 0 (nowcasts) and no implicit publication lag of the target variable

[26]In the expanding window approach, the total number of training, validation and test steps is determined and limited by the number of observations of the data.

[27]Note that the above described procedure is valid for predictions of horizon 0 (nowcasting) and, in particular, only if the target variable does not show an (implicit) publication lag. If – as in our application – the target variable is only available with a lag (see Section 3), for a given validation point, only those preceding observations for which the target would have already been feasible must serve as the training set. If a red/blue dot in Figure C.1 would, for example, represent a February observation of our monthly GDP measure, the last yellow point before the red/blue dot would not be 'feasible' and must not be used. Analogously, only the prediction performances in validation observations prior to a test point that would have been 'feasible' must be evaluated for model selection.

D Additional results figures and tables

D.1 Performance of payment compared to direct benchmark models

		Payment model CARD-SIC			
		RMSEs	RMSEs relative to RMSEs of payment model SIC / percental relative performance	RMSEs relative to RMSEs of 'long' direct benchmark model / percental relative performance	RMSEs relative to RMSEs of 'short' direct benchmark model / percental relative performance
Non-crisis ('short' sample of payment/direct benchmark models)	h = -3	0.00629	0.96 / 4%	0.97 / 3%	0.93 / 7%
	h = -2	0.00722	0.90 / 11%	0.94 / 7%	0.89 / 12%
	h = -1	0.00745	0.97 / 3%	1.03 / -3%	0.96 / 5%
	h = 0	0.00881	0.98 / 2%	1.14 / -14%	0.99 / 1%
	h = 1	0.00966	1.03 / -3%	1.19 / -19%	1.04 / -4%
	h = 2	0.01031	0.96 / 4%	1.25 / -25%	1.06 / -6%
	h = 3	0.01127	1.04 / -4%	1.32 / -32%	1.13 / -13%
COVID-19 crisis (sample of payment/direct benchmark models)	h = -3	0.02339	0.91 / 10%	0.86 / 16%	0.71 / 41%
	h = -2	0.02826	0.88 / 14%	0.87 / 15%	0.78 / 28%
	h = -1	0.03187	0.90 / 11%	0.91 / 10%	0.77 / 29%
	h = 0	0.03551	0.88 / 14%	0.84 / 19%	0.69 / 45%
	h = 1	0.03811	0.94 / 7%	0.88 / 13%	0.59 / 68%
	h = 2	0.04771	1.06 / -6%	1.09 / -9%	0.70 / 43%
	h = 3	0.05303	1.15 / -15%	1.19 / -19%	0.75 / 33%

Table D.1: Pseudo real-time out-of-sample performance of the payment model CARD-SIC

		Payment model SIC		
		RMSEs	RMSEs relative to RMSEs of 'long' direct benchmark model / percental relative performance	RMSEs relative to RMSEs of 'short' direct benchmark model / percental relative performance
Non-crisis ('short' sample of payment/direct benchmark models)	h = -3	0.00654	1.01 / -1%	0.97 / 3%
	h = -2	0.00800	1.04 / -4%	0.99 / 1%
	h = -1	0.00769	1.06 / -6%	0.99 / 1%
	h = 0	0.00898	1.16 / -16%	1.01 / -1%
	h = 1	0.00938	1.16 / -16%	1.01 / -1%
	h = 2	0.01077	1.30 / -30%	1.11 / -11%
	h = 3	0.01089	1.28 / -28%	1.09 / -9%
COVID-19 crisis (sample of payment/direct benchmark models)	h = -3	0.02584	0.96 / 5%	0.78 / 28%
	h = -2	0.03214	0.99 / 1%	0.89 / 13%
	h = -1	0.03545	1.01 / -1%	0.86 / 16%
	h = 0	0.04051	0.96 / 4%	0.78 / 27%
	h = 1	0.04070	0.94 / 6%	0.63 / 58%
	h = 2	0.04516	1.03 / -3%	0.66 / 51%
	h = 3	0.04631	1.04 / -4%	0.66 / 53%
Great financial crisis (sample of SIC/ 'long' direct benchmark model)	h = -3	0.01024	1.16 / -16%	- / -
	h = -2	0.01296	0.92 / 8%	- / -
	h = -1	0.01495	0.65 / 55%	- / -
	h = 0	0.02458	0.60 / 65%	- / -
	h = 1	0.02627	0.45 / 123%	- / -
	h = 2	0.03428	0.38 / 161%	- / -
	h = 3	0.03840	0.34 / 195%	- / -

Table D.2: Pseudo real-time out-of-sample performance of the payment model SIC

		Direct benchmark models		
		RMSEs of 'long' direct benchmark model	RMSEs of 'short' direct benchmark model	RMSEs relative to RMSEs of 'short' direct benchmark model / percental relative performance
Non-crisis ('short' sample of payment/direct benchmark models)	h = -3	0.00647	0.00674	0.96 / 4%
	h = -2	0.00771	0.00808	0.95 / 5%
	h = -1	0.00723	0.00779	0.93 / 8%
	h = 0	0.00772	0.0089	0.87 / 15%
	h = 1	0.00809	0.00931	0.87 / 15%
	h = 2	0.00826	0.00974	0.85 / 18%
	h = 3	0.00853	0.00998	0.86 / 17%
COVID-19 crisis (sample of payment/direct benchmark models)	h = -3	0.02705	0.03298	0.82 / 22%
	h = -2	0.03251	0.03628	0.90 / 12%
	h = -1	0.03517	0.0412	0.85 / 17%
	h = 0	0.04222	0.05162	0.82 / 22%
	h = 1	0.04319	0.06415	0.67 / 49%
	h = 2	0.04374	0.06817	0.64 / 56%
	h = 3	0.04442	0.07067	0.63 / 59%
Great financial crisis (sample of SIC/'long' direct benchmark model)	h = -3	0.00886	-	- / -
	h = -2	0.01404	-	- / -
	h = -1	0.02317	-	- / -
	h = 0	0.04065	-	- / -
	h = 1	0.05859	-	- / -
	h = 2	0.08964	-	- / -
	h = 3	0.11343	-	- / -

Table D.3: Pseudo real-time out-of-sample performance of the direct benchmark models

D.2 Performance of quarterly payment models compared to quarterly direct benchmark model (robustness check)

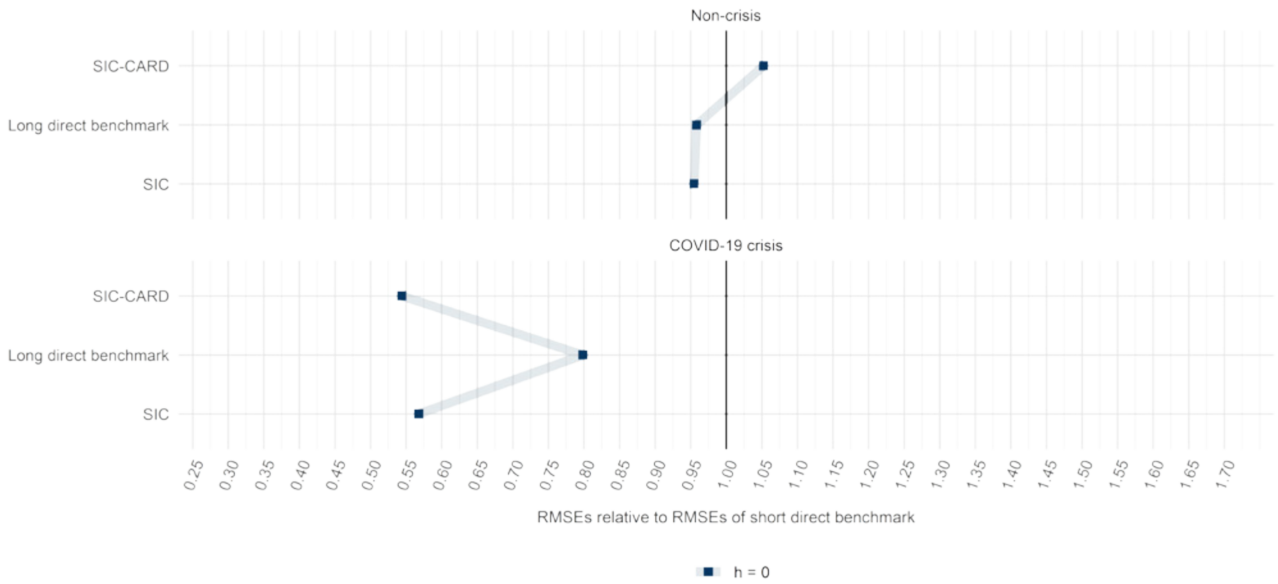

Figure D.1: The real-time out-of-sample RMSEs of the quarterly prediction exercise (robustness check) confirm the state-dependent value of payments data for nowcasting economic activity: the payment models' out-of-sample nowcasts are superior to those of the direct benchmark models in times of turmoil (COVID-19 crisis) but not (substantially) in 'normal' non-crisis times.

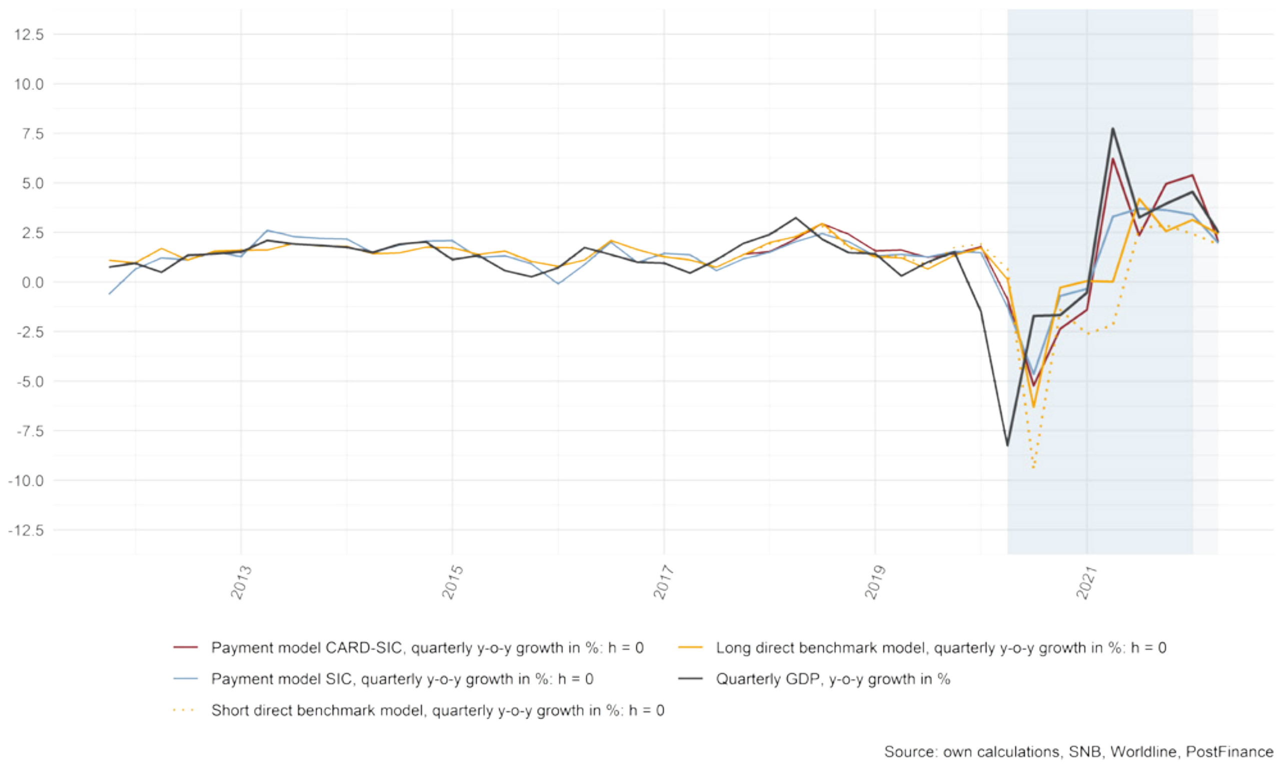

Figure D.2: Payments data are particularly valuable during times of crisis: the quarterly nowcasts of the payment models outperform the direct benchmark models during the pandemic (shaded area).

		Payment model CARD-SIC, quarterly									
		RMSEs	RMSEs relative to RMSEs of payment model SIC / percental relative performance			RMSEs relative to RMSEs of 'long' direct benchmark model / percental relative performance			RMSEs relative to RMSEs of 'short' direct benchmark model / percental relative performance		
Non-crisis	h = 0	0.01272	1.10	/	-10%	1.10	/	-10%	1.05	/	-0.05
COVID-19 crisis	h = 0	0.02854	0.96	/	4%	0.68	/	47%	0.54	/	0.84

Table D.4: Pseudo real-time out-of-sample performance of the quarterly payment model CARD-SIC

		Payment model SIC, quarterly						
		RMSEs	RMSEs relative to RMSEs of 'long' direct benchmark model / percental relative performance			RMSEs relative to RMSEs of 'short' direct benchmark model / percental relative performance		
Non-crisis	h = 0	0.01153	1.00	/	0%	0.95	/	5%
COVID-19 crisis	h = 0	0.02980	0.71	/	41%	0.57	/	76%

Table D.5: Pseudo real-time out-of-sample performance of the quarterly payment model SIC

		Direct benchmark models, quarterly				
		RMSEs of 'long' direct benchmark model	RMSEs of 'short' direct benchmark model	RMSEs relative to RMSEs of 'short' direct benchmark model / percental relative performance		
Non-crisis	h = 0	0.01158	0.01	0.96	/	4%
COVID-19 crisis	h = 0	0.04192	0.05	0.80	/	25%

Table D.6: Pseudo real-time out-of-sample performance of the quarterly direct benchmark models

D.3 Performance of payment models compared to practical benchmark

		Payment model CARD-SIC				
		RMSEs	RMSEs with latest vintage target	RMSEs relative to RMSEs with latest vintage target	RMSEs relative to RMSEs of practical benchmark model / percental relative performance	Latest vintage target: RMSEs relative to RMSEs of practical benchmark model / percental relative performance
Non-crisis (sample of practical benchmark)	h = -3	0.00629	0.00605	1.04 / -4%	1.07 / -7%	0.89 / 12%
	h = -2	0.00722	0.00635	1.14 / -14%	1.04 / -4%	0.90 / 11%
	h = -1	0.00745	0.00690	1.08 / -8%	1.12 / -12%	0.93 / 8%
	h = 0	0.00881	0.00705	1.25 / -25%	1.29 / -29%	0.97 / 3%
	h = 1	0.00966	0.00809	1.19 / -19%	1.39 / -39%	1.13 / -13%
	h = 2	0.01031	0.00841	1.23 / -23%	1.48 / -48%	1.20 / -20%
	h = 3	0.01097	0.00911	1.20 / -20%	1.45 / -45%	1.27 / -27%
COVID-19 crisis (sample of practical benchmark)	h = -3	0.02339	0.02301	1.02 / -2%	0.99 / 1%	0.96 / 4%
	h = -2	0.02826	0.02705	1.04 / -4%	1.00 / 0%	0.95 / 6%
	h = -1	0.03187	0.02863	1.11 / -11%	0.99 / 1%	0.93 / 7%
	h = 0	0.03551	0.03152	1.13 / -13%	0.97 / 3%	0.92 / 9%
	h = 1	0.03811	0.03324	1.15 / -15%	0.95 / 5%	0.90 / 11%
	h = 2	0.04771	0.04465	1.07 / -7%	1.09 / -9%	1.11 / -11%
	h = 3	0.05282	0.05035	1.05 / -5%	1.07 / -7%	1.14 / -14%

Table D.7: Pseudo real-time and latest vintage out-of-sample performance of the payment model CARD-SIC

		Payment model SIC				
		RMSEs	RMSEs with latest vintage target	RMSEs relative to RMSEs with latest vintage target	RMSEs relative to RMSEs of practical benchmark model / percental relative performance	Latest vintage target: RMSEs relative to RMSEs of practical benchmark model / percental relative performance
Non-crisis (sample of practical benchmark)	h = -3	0.00654	0.00735	0.89 / 12%	1.11 / -11%	1.08 / -8%
	h = -2	0.00793	0.00768	1.03 / -3%	1.14 / -14%	1.09 / -9%
	h = -1	0.00761	0.00780	0.98 / 2%	1.14 / -14%	1.05 / -5%
	h = 0	0.00864	0.00890	0.97 / 3%	1.26 / -26%	1.23 / -23%
	h = 1	0.00923	0.00994	0.93 / 8%	1.33 / -33%	1.39 / -39%
	h = 2	0.0104	0.01179	0.88 / 13%	1.49 / -49%	1.69 / -69%
	h = 3	0.00987	0.01196	0.83 / 21%	1.30 / -30%	1.66 / -66%
COVID-19 crisis (sample of practical benchmark)	h = -3	0.02584	0.02712	0.95 / 5%	1.10 / -10%	1.13 / -13%
	h = -2	0.03214	0.03399	0.95 / 6%	1.14 / -14%	1.19 / -19%
	h = -1	0.03545	0.0347	1.02 / -2%	1.10 / -10%	1.13 / -13%
	h = 0	0.04051	0.03954	1.02 / -2%	1.10 / -10%	1.16 / -16%
	h = 1	0.04070	0.03865	1.05 / -5%	1.02 / -2%	1.05 / -5%
	h = 2	0.04516	0.04283	1.05 / -5%	1.03 / -3%	1.06 / -6%
	h = 3	0.04760	0.04576	1.04 / -4%	0.97 / 3%	1.03 / -3%

Table D.8: Pseudo real-time and latest vintage out-of-sample performance of the payment model SIC

		Practical benchmark model		
		RMSEs	RMSEs with latest vintage target	RMSEs relative to RMSEs with latest vintage target
Non-crisis (sample of practical benchmark)	h = -3	0.00588	0.00678	0.87 / 15%
	h = -2	0.00695	0.00702	0.99 / 1%
	h = -1	0.00667	0.00743	0.90 / 11%
	h = 0	0.00683	0.00724	0.94 / 6%
	h = 1	0.00694	0.00717	0.97 / 3%
	h = 2	0.00696	0.00699	1.00 / 0%
	h = 3	0.00757	0.00720	1.05 / -5%
COVID-19 crisis (sample of practical benchmark)	h = -3	0.02354	0.02404	0.98 / 2%
	h = -2	0.02821	0.02858	0.99 / 1%
	h = -1	0.03228	0.03078	1.05 / -5%
	h = 0	0.03673	0.03422	1.07 / -7%
	h = 1	0.04009	0.03677	1.09 / -9%
	h = 2	0.04380	0.04028	1.09 / -9%
	h = 3	0.04924	0.04436	1.11 / -11%

Table D.9: Pseudo real-time and latest vintage out-of-sample performance of the practical benchmark

Recent SNB Working Papers

SCHWEIZERISCHE NATIONALBANK
BANQUE NATIONALE SUISSE
BANCA NAZIONALE SVIZZERA
BANCA NAZIUNALA SVIZRA
SWISS NATIONAL BANK