ECONOMIC MEMO

MICRO DATA SCIENCE

DANMARKS NATIONALBANK

23 MARCH 2021 — NO. 3

A factor model approach to nowcasting Danish GDP

Adrian Michael Bay Schmith
Economist
ECONOMICS AND MONETARY POLICY
ambs@nationalbanken.dk

Erik Grenestam
Data Scientist
FINANCIAL STATISTICS
egr@nationalbanken.dk

A factor model approach to nowcasting Danish GDP

Abstract

Nowcasting models help support short-term assessment of the economy. In light of the outbreak of the COVID-19 pandemic, we propose a dynamic factor nowcast model approach in order to link soft and hard indicators of the Danish economy to GDP growth.

Our nowcast model provides timely signals during periods of volatile growth. In 2020, our model's forecast error was 62 per cent lower compared to a simple mean growth rate prediction.

We use new high-frequency, real-time indicators to forecast traditional monthly indicators in order to further improve the timeliness of the model.

Up-to-date knowledge of the current economic situation is important in order to inform economic policy. Especially in times of large business cycle fluctuations, timely information is key to the identification of appropriate policy measures. Very short-term forecasting of aggregate economic outcomes, also known as nowcasting, can be a key tool to gain such knowledge.

An increasing number of soft and hard indicators are available for surveilling short-term business cycle developments. A nowcast model is a statistical model that uses such timely indicators to produce a near-term forecast of aggregate outcomes. The value added of a nowcast model is advance knowledge of aggregate economic indicators such as the gross domestic product (GDP).

In this memo, we describe and evaluate a GDP nowcasting model used by Danmarks Nationalbank during the economic crisis related to the outbreak of COVID-19. The model is commonly known as a dynamic factor model.

First, we show that while the model generally does not outperform a simple mean growth rate prediction, it provided timely and more accurate nowcasts during the COVID-19 crisis. Second, we make use of new high-frequency indicators to further improve the timeliness of the model.

An urgent need for timely information as the pandemic surged

The outbreak of the COVID-19 pandemic in the spring of 2020 called for swift and targeted policy responses across the world, and inevitably also in Denmark. Quick updates were needed to understand the economic impact of the pandemic and the subsequent lockdown. This led both statistical agencies in Denmark, including Statistics Denmark and Danmarks Nationalbank, and private companies to develop novel high-frequency weekly and even daily indicators.

To meet the urgent demand for linking both new and existing indicators to aggregate economic activity, we chose a well-documented and tested model that is able to take in information from indicators as they arrive and link them to GDP growth. Specifically, we build our nowcasting model on a Dynamic Factor Model (DFM) published by the New York Federal Reserve (Bok et al., 2017). See Box 1 for a technical description of the model. [1]

Data covers the main aspects of the macro economy

We structure indicators relevant for the Danish economy into groups that cover the major demand components of GDP (private consumption, foreign trade and production), key markets of the macro economy (the labour market, the financial market and the housing market) and price developments.

Our choice of data takes inspiration from a vast literature on nowcasting GDP (see e.g. Aastveit and Trovik, 2012; Bok et al., 2017; Chernis and Sekkel, 2017). We rely predominantly on monthly indicators, although some quarterly indicators are also considered. Previous work suggests that the degree of relevant information from different economic indicators can be very context-specific. In Norway, financial indicators have been shown to be important predictors of GDP (Aastveit & Trovik, 2012), whereas Rusnák (2013) shows that financial indicators are redundant in nowcasting the Czech Republic GDP.

As noted by Ankargren & Lindholm (2021), the consensus regarding data for nowcasting has shifted from including a very large number of disaggregated series to fewer aggregate numbers, while maintaining coverage of all GDP components. We select 34 series from a large number of candidates by testing how different combinations of variables have performed historically. We also take into consideration that all series have a sufficient historical coverage and are updated in a timely manner. Our choice of variables is presented in Table 1.

Because Denmark is a small open economy, capturing the development of imports and exports is an essential component of a GDP forecast. However, nowcasting foreign trade figures is notoriously difficult (Higgins, 2014). We therefore include a large number of foreign trade indicators, with a focus on the largest export markets for Denmark.

To a large extent, the data has similarities to data previously used for short-term forecasting in Nationalbanken, cf. Jørgensen, Storgaard and Sørensen (2011). However, our framework allows for an easier implementation of large data sets. Further, our model is also able to handle the so-called "ragged edge" problem. This problem arises because data is not all updated at the same time, leading to different patterns of missing data for the quarter we want to forecast. In order to facilitate a continuously updated nowcast, our model is able to make use of all available data to forecast the missing figures.

Clear relevance in periods of volatile growth

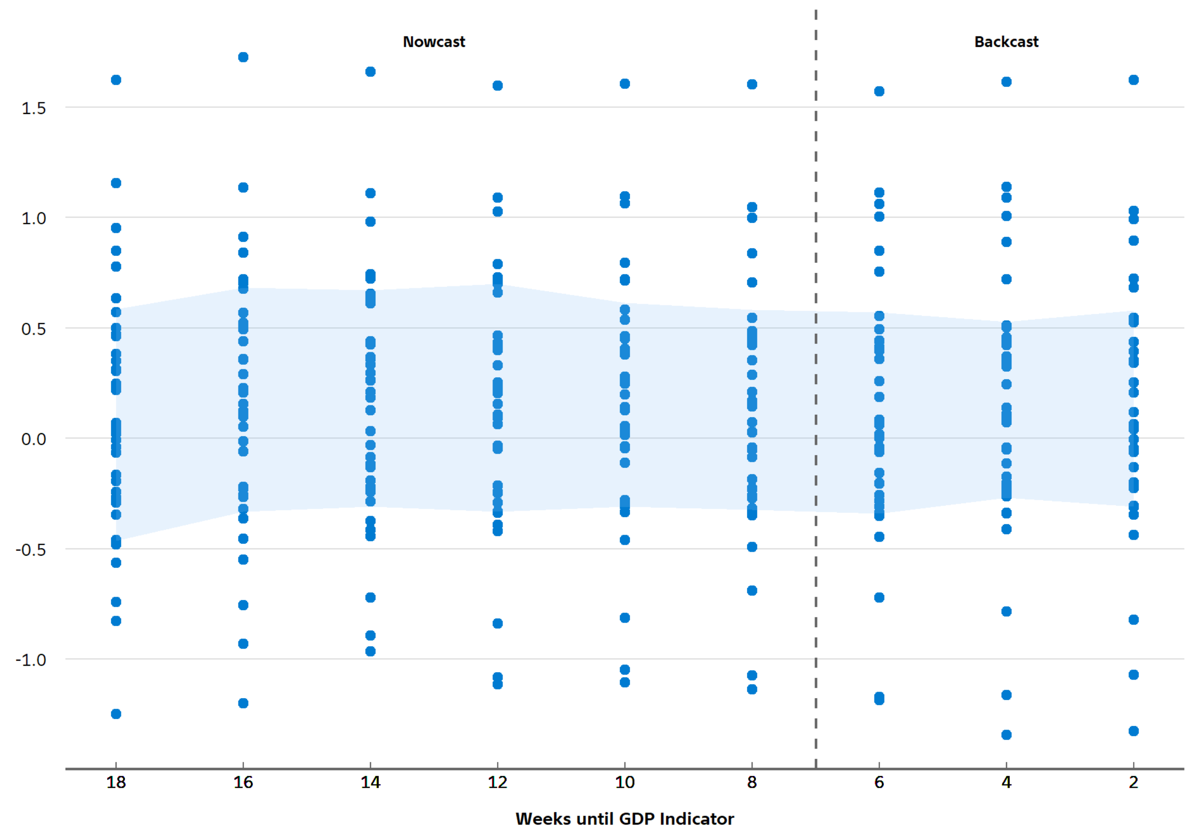
We test the performance of our model in the period from 2010Q3 to 2019Q4. We exclude 2020 from our main evaluation to minimise the influence of extreme observations. Before each quarter, we make nine vintage data sets over an 18-week period prior to the first release of the official GDP number. Each vintage thus represents the data that would have been available to the forecaster for every second week leading up to the first official release of GDP. [2]

[1] The application to the Danish setting remains work in progress, and this memo reflects a snapshot of the state of the model.

[2] As some data is revised after its initial release, and we do not have access to historical unrevised data but are limited to the latest available data, there will be differences between our vintages and the data that was available in real time. The literature refers to the construction of such vintages as "pseudo real-time data".

Forecast errors generally do not improve over the quarter

Chart 1



Note: Forecast errors based on out-of-sample forecasts from 2010Q3 to 2019Q4. Each dot represents a forecast made in a given week prior to the official GDP Indicator is published. We use the latest available GDP numbers, which are subject to revision up to three years after initial release. The blue area indicates the 84th and 16th percentile of forecast errors, corresponding to +/- 1 standard deviation assuming errors are normally distributed.

However, GDP numbers are continuously revised for up to three years after initial release. We choose to target the latest available number when testing our model's performance, as this number is the most accurate measure of aggregate economic activity. Our estimation sample is recursively expanding each quarter. We estimate the model once every quarter using the first data vintage and then evaluate all subsequent data releases keeping model parameters fixed in order to examine the role of news during the quarter.

Chart 1 presents our evaluation of historical model performance in terms of the difference between our forecast and the official GDP numbers produced by DST, the forecast error. We see that the forecast error is basically unchanged from the initial forecast made in the beginning of the quarter to the forecast made about four weeks after the quarter ends (two weeks before the first official GDP number is released). While this is a sobering result, it is not uncommon for a DFM to show no or marginal improvements in performance over the quarter (Bok et al., 2017; Ankargren & Lindholm, 2021).

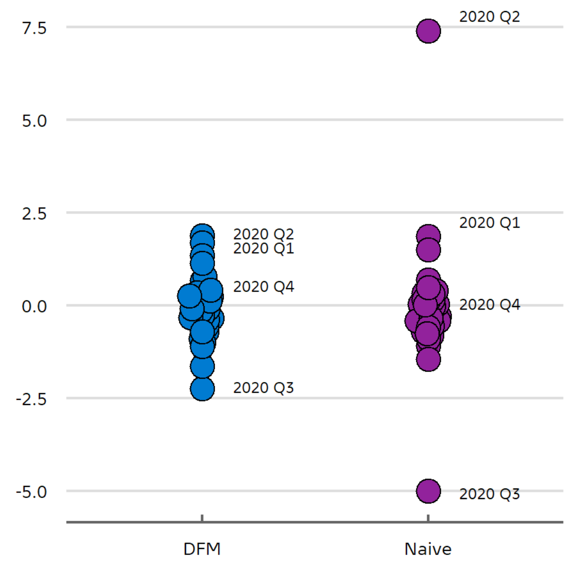
While we target the latest available figures, it is natural that nowcasts are compared to the first officially released GDP figure. Revisions to GDP after the first release can be substantial. As an example, the root-mean-squared forecast error (RMSFE) of using the first official growth figure from Statistics Denmark as an estimate for the latest revised figure is about 0.5 for our evalution period, only slightly below the RMSFE of 0.6 that our model produces.[3] This reflects the difficulty of producing an accurate nowcast, even after the quarter ends.

While the nowcast does not improve over the quarter in general, the ability to quickly incorporate changes in economic indicators has proven useful during the COVID-19 pandemic. Chart 2 presents the forecast errors of the model, including 2020. To put our forecast error in context, we use a benchmark prediction that the growth rate for the current quarter will be the mean growth rate of past quarters.[4] We see that during 2020, the DFM outperformed this simple benchmark, with RMSFE for 2020 being about 62 per cent lower. However, when excluding 2020 the models have had similar historical accuracy. This illustrates that the DFM is most relevant during times of volatile growth, which is also likely to be when the interest in advance knowledge of GDP is at its highest.

In addition to providing a nowcast of GDP, the DFM can also inform on the impact that each new data point has on the nowcast. In Chart 3, we show how new data affected the nowcast during the first quarter of 2020 over the final weeks of the quarter. The impact of new data is categorised in order to ease interpretation. This representation of model results is currently released weekly on the website of Danmarks Nationalbank alongside another nowcast model (Pedersen, 2021).

The nowcast model is most useful when changes to GDP are large

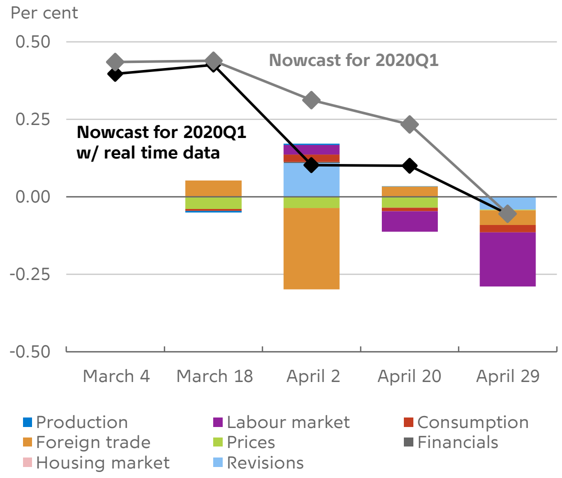
Chart 2



Note: Comparison of forecast errors from our DFM (blue) versus a naive prediction of the past mean growth rate (purple). The DFM forecast errors are based on estimates from X weeks away from the release of GDP.

We can calculate the impact of each piece of news on the nowcast

Chart 3



Note: Evolution of the nowcast during the end of 2020Q1 with and without real-time indicators. The bars represent of impact of new data on the nowcast (in grey). The two nowcasts converge when all data for the quarter in the extended series has been released.

3 This does not imply that our forecasts are very close to the initial growth figure. When we target the initial growth figure, the DFM performs slightly better.

4 Such a naive prediction outperforms any auto-regressive specification when including 2020.

Extending the model with real-time data to enhance timeliness

Any nowcast model depends on the timeliness with which economic indicators are available. Several of the key indicators in our model are published with a substantial lag, often several weeks past the end of the reference period.

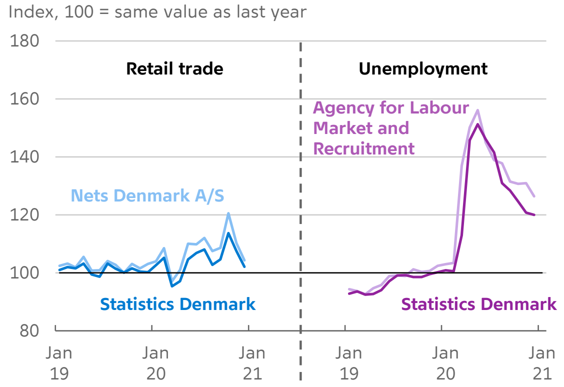
The development of new high-frequency indicators updated on a weekly or even daily basis has seen a surge since the outbreak of the pandemic. Danmarks Nationalbank has monitored several high-frequency indicators closely during the pandemic in order to support the assessment of the near-term outlook for the Danish economy (Danmarks Nationalbank, 2020).

To further enhance the ability to quickly pick up changes to the economic outlook, we investigate the possibility of including such indicators in the dynamic factor model framework. Due to the novelty of many real-time indicators, a limiting factor is the amount of historical data available. This poses an estimation challenge and prevents us from including real-time indicators directly in the DFM.

Instead, we implement a simple two-step process, where high-frequency indicators are first used for a short-term forecast of some monthly indicators in the existing model. That is, we use daily data to predict the first missing value of the monthly indicator. Specifically, we extend unemployment data using daily data on the number of people filing for unemployment insurance benefits and we extend the retail trade index using daily payment systems data. Chart 4 shows the scope of the extensions, while Box 2 provides technical details on how we set up statistical models to do so. In the second step, we then add the prediction-extended monthly series to the DFM and hence provide the model with more timely information.

Due to the short sample of the two real-time indicators, we are unable to fully evaluate their impact on the performance of the DFM. However, our statistical models in the first step are able to generate accurate predictions which we expect to benefit the model. For example, the extended model was quicker at capturing the downturn in Q1 2020 following the outbreak of COVID-19 (see Chart 4).

High-frequency indicators correlate with monthly series — Chart 4



Anm.: The chart shows daily data from card payments from Nets and registrations for unemployment from the Agency for Labour Market and Recruitment compiled to a monthly level. This is compared to monthly series for retail trade and gross unemployment from Statistics Denmark. Indices are not seasonally adjusted.

Kilde: Nets Denmark A/S, the Danish Agency for Labour Market and Recruitment, Statistics Denmark and own calculations.

The work on extending the dynamic factor model with real-time indicators is still in progress. At present, work on including real-time indicators for truck traffic, flight traffic and Google search patterns is ongoing. As new indicators are made available, and the sample size of existing real-time indicators increases, there is scope for improving their contribution to the nowcast model.

Variable specification

Table 1

Category	Variable	Publication lag (days)	Frequency	Transform	Typical factor loadings
Production	Gross Domestic Product (GDP)	45	Q	QPC	-0.05, 0.01
	PMI	1	M	FD	-0.07, -0.13
	Business sentiment	-2	M	YPC	-0.25, 0.04
Labour market	Gross unemployment	30	M	FD	0.28, -0.25
	Wage-earner employment	52	M	FD	-0.22, 0.21
	Job ads	7	M	FD	-0.20, -0.07
	Unemployment rate	30	M	FD	0.22, -0.19
Private consumption	Consumer confidence	-7	M	FD	-0.16, -0.20
	Retail trade	25	M	PC	-0.07, -0.34
	Auto sales	10	M	MPC	-0.08, -0.02
Prices	Core inflation	10	M	MPC	0.03, 0.10
	Producer prices	15	M	MPC	-0.08, 0.04
	Export prices	15	M	MPC	-0.06, 0.05
	Oil prices	-	M	MPC	-0.10, -0.35
Financial conditions	Short-term mortgage bond rate	-	M	FD	-0.04, 0.25
	OMX Copenhagen 25 index	-	M	MPC	-0.07, -0.21
Housing market	Construction permits	45-135	M	YPC	-0.04, 0.13

Notes: Publication lag refers to the number of days relative to the end of the reference period before data for the period is released. Transformation codes: QPC: quarterly percentage change, MPC: monthly percentage change, YPC: annual percentage change, FD: first difference. Typical factor loading refers to a single model estimation using our preferred two-factor specification, larger magnitudes indicate a higher correlation with the factors. Quarterly variables' load on a linear combination of the two factors (see equation 3 in Box 1).
Source: Macrobond, Statistics Denmark and JobIndex.

Variable specification (continued)

Table 2

Category	Variable	Publication lag (days)	Frequency	Transform	Typical factor loadings
Foreign trade	Overnight stays at hotels	40	M	YPC	-0.05, 0.16
	Export-weighted manufacturing PMI	-10	M	FD	-0.22, -0.09
	Goods export	40	M	MPC	-0.11, 0.22
	Goods import	40	M	MPC	-0.16, 0.22
	Services export	40	M	MPC	-0.10, 0.27
	Euro area GDP	30	Q	QPC	-0.04, -0.05
	Euro area unemployment rate	30	M	FD	0.03, -0.30
	Euro area HICP	-1	M	MPC	0.02, 0.13
	EU27 retail trade	40	M	MPC	-0.25, -0.29
	EU27 economic sentiment	-5	M	FD	-0.36, -0.02
	German IFO indicator	-5	M	FD	-0.28, -0.13
	Euro area export orders	-1	M	MPC	0.02, 0.00
	Euro area production expectations	-2	M	MPC	0.05, 0.00
	U.S. GDP	30	Q	QPC	-0.05, -0.05
	Euro area consumer sentiment	-5	M	FD	-0.29, -0.12
	Euro area service sentiment	-5	M	FD	-0.33, 0.06

Source: Macrobond, Statistics Denmark and JobIndex.

An established framework for nowcasting GDP

Box 1

To nowcast Danish GDP at a quarterly frequency, we apply a dynamic factor model similar to Bok et al. (2017). The model is estimated using maximum likelihood and allows for arbitrary patterns of missing data, which allows us to make use of new data points as they are released during the quarter.

The dynamic factor framework is a parsimonious way of modelling data with a large number of distinct time series. Our starting point is a panel data set consisting of N time series that jointly describe the state of the economy ($y_{1t}, \ldots y_{Nt}$). We assume that our data is driven by common unobservable factors (we can think of a factor as describing the underlying state of the economy), along with idiosyncratic processes unique to each time series. Our model can be described by the following equation:

(1) $$y_{it} = \lambda_i f_t + e_{it}$$

where f_t is a vector consisting of a small number of common unobservable factors (we set the number of factors to 2). The dynamics of the model are assumed to be generated by a first order auto-regressive process according to:

(2) $$f_t = a f_{t-1} + u_t$$

where u_t is an iid error term assumed to be normally distributed. The majority of our variables are updated on a monthly schedule, but some (notably GDP) are only observed quarterly. We address this frequency mismatch using the approach by Mariano & Murasawa (2003). The quarterly growth rate of GDP can be approximately represented as a sum of monthly growth rates. We treat the unobserved monthly growth rates as factors to be estimated.

(3) $$y_t^Q \approx y_t^M + 2y_{t-1}^M + 3y_{t-2}^M + 2y_{t-4}^M + y_{t-5}^M$$

An important problem facing any macroeconomic forecast model is that news about the economy is not released at a single point in time during each month or quarter, but rather as a steady trickle. This gives rise to what is sometimes called a ragged edge in the data, as data for the current quarter will typically have missing values. The estimation procedure that we use is an expectation maximisation algorithm that extracts a signal from any data that has been released even if other data for the current period is missing. This also allows us to gauge the information content and the level of surprise associated with each new data point, and how it affects the GDP nowcast.

To get the initial factor estimate, we take the principal components from our data. Then, the estimation procedure iterates on two steps. First, we estimate equation (1) using OLS. Second, we re-estimate the factors using the Kalman filter. We iterate these two steps until the log likelihood function converges. For a complete description of the model and estimation procedures, see Banbura & Modugno et al. (2014) and de Valk et al. (2019).

Nowcasting unemployment and retail trade using real-time data

Box 2

Retail trade and unemployment are economic indicators that provide important information when forecasting GDP growth. The monthly series published by Statistics Denmark are subject to a three- and eight-week publication lag for retail trade and unemployment, respectively. To enhance the timeliness of our nowcasting of GDP growth, we implement one-month-ahead nowcasts of retail trade and unemployment using daily data available in real time.

For unemployment, the Danish Agency for Labour Market and Recruitment (STAR) provides us with daily updates on the number of people entering into or leaving various unemployment insurance benefit programmes. While the underlying data is similar to the data used by Statistics Denmark's register-based unemployment statistics, there are discrepancies due to, among other things, the recalculation into full-time equivalents undertaken by Statistics Denmark.

For retail trade, we have access to daily payment data from payment cards issued in Denmark and processed by Nets Denmark A/S.[1] The sum of all payments made to retailers is a natural predictor of the retail trade turnover index.

We use a mixed-data sampling (MIDAS) model to account for the different publication frequencies. The model can be written as

(4) $$\Delta y_t = \alpha + \gamma \Delta y_{t-1} + \beta B\left(L^{1/m};\theta\right) x_t^{(m)} + \varepsilon_t$$

where Δy_t is the change in the dependent variable and $x_t^{(m)}$ is the high-frequency predictor, where m denotes the number of high-frequency observations per month. $B\left(L^{1/m};\theta\right)$ is a function determining the lag coefficients for each of the lags in $x_t^{(m)}$. $L^{1/m}$ is a lag operator such that $L^{k/m} x_t^{(m)} = x_{t-k/m}^{(m)}$. To clarify this notation, if we assume daily data and a 30-day month we set $m = 30$ and we have that $x_{t-7/30}^{(30)}$ is the value of x seven days prior to the end of month t. The key feature of a MIDAS model as opposed to a regular distributed lag regression is the function $B\left(L^{1/m};\theta\right)$. To maintain a parsimonious specification robust to autocorrelation in x, we assume that the lag coefficients are determined by a small number of parameters independent of the number of lags (θ). We use the so-called Almon polynomial and set $\theta \in \mathbb{R}^2$. To account for intra-week seasonality, we aggregate the daily data to full weeks.

For unemployment, we set $m = 6$, starting from seven days before the first of each month and six weeks forward. This month-to-month overlap helps to account for potential reporting irregularities for individuals who become unemployed on the last day of the month. We use the non-seasonally adjusted series and take the first difference in monthly unemployment. The predicted change is then transformed back to levels and seasonally adjusted using the adjustment factor from the same month in the previous year. Using an expanded window out-of-sample evaluation for the past three years, our model yields an RMSE that is 13 per cent lower compared to a simple auto-regressive model with one lag, AR(1).

For retail trade, we set $m = 5$, starting on the first of each month and counting forward five weeks. We use the non-seasonally adjusted chain-linked quantity index for retail trade. To account for seasonality in the retail trade data as well as the retail trade index, we transform both variables to 12-month differences. The predicted change is then transformed back to levels and seasonally adjusted using the adjustment factor from the same month in the previous year. Evaluating model performance is made difficult by the fact that we only have three years of data. Using an expanded window out-of-sample evaluation for the past 12 months, our model yields an RMSE that is 61 per cent lower compared to a simple AR(1) model.

1. This represents the vast majority of card transactions undertaken by Danes.

References

Aastveit, K.A., Trovik, T. (2012), Nowcasting Norwegian GDP: the role of asset prices in a small open economy. Empirical Economics 42, 95-119.

Ankargren, S. and Lindholm, U. (2021), Nowcasting Swedish GDP Growth. National Institute of Economic Research Working Paper 154.

Banbura, M., Giannone, D. and Reichlin, L. (2010), Nowcasting. ECB Working Paper No. 1275.

Banbura, M. and Modugno, M. (2014), Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. J. Appl. Econ., 29: 133-160.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A. and Tambalotti, A. (2017), Macroeconomic Nowcasting and Forecasting with Big Data. Federal Reserve Bank of New York Staff Reports 830.

Chernis, T. and Sekkel, R. (2017), A dynamic factor model for nowcasting Canadian GDP growth. Empirical Economics 53(4).

Danmarks Nationalbank (2020), Danish and international economy hit by pandemic. Danmarks Nationalbank Analyse 2020:4.

de Valk, S., de Mattos, D., Ferreira, P. (2019), Nowcasting: An R Package for Predicting Economic Variables Using Dynamic Factor Models. The R Journal (2019) 11:1, pages 230-244.

Higgins, P. (2014), GDPNow: A model for GDP Nowcasting. FRB Atlanta Working Paper 2014-7.

Hvid Jørgensen, E., Storgaard, P.E. and Sørensen, J., Værktøjer til vurdering af konjunktursituationen, Danmarks Nationalbank Quarterly Review, 4th quarter 2011.

Pedersen, J. (2021), Detecting turning-points in the Danish economy in real-time. Danmarks Nationalbank Working Paper 172.

Rusnák, M. (2013), Nowcasting Czech GDP in real time. Czech National Bank Working Paper 2013:6.

Schmith, A., Sørensen, R. and Werner, M. (2020), Using input-output methods and high frequency indicators in assessment of GDP in 2020. Danmarks Nationalbank Economic Memo 202:12.

MICRO DATA SCIENCE

Data in new ways

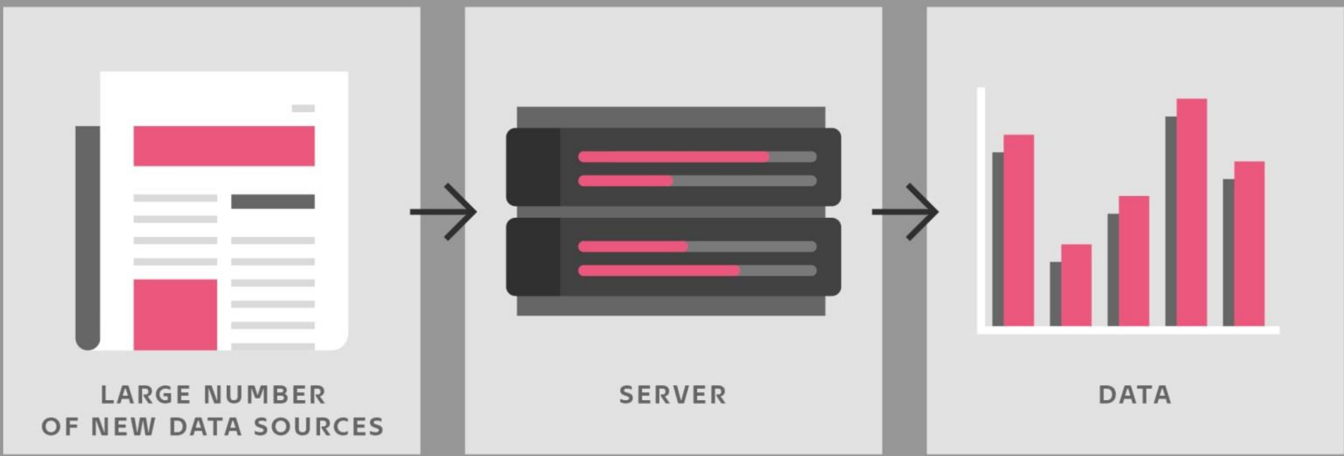
Data volumes have grown exponentially. By 2025, an estimated 450 exabytes of data will be created each day.

This is equivalent to hundreds of millions of personal computers being filled with data on a daily basis. The vast volumes of data are highly diverse, but new and sophisticated methods enable analysis of this data in new and more efficient ways.

New data types and new data collection methods may be used in various contexts in Danmarks Nationalbank's ongoing work.

In order to acquire more knowledge and a better basis for assessing the Danish economy, Danmarks Nationalbank focuses on new data types and methods in a series of publications of which this Economic Memo is one.

New data creates new knowledge



PUBLICATIONS

NEWS

News offers a quick and accessible insight into an Analysis, an Economic Memo, a Working Paper or a Report from Danmarks Nationalbank. News is published continuously.

ANALYSIS

Analysis from Danmarks Nationalbank focuses on economic and financial matter. Some of the analyses are published with a regular frequency e.g. *Outlook for the Danish economy and Financial stability*. Other analyses are published continuously.

REPORT

Report comprises recurring reports and reviews of the functioning of Danmarks Nationalbank. For instance Report includes the *Annual report* and the annual publication *Danish government borrowing and debt*.

ECONOMIC MEMO

Economic Memo is a cross between Analysis and Working Paper and it often shows the ongoing study of the authors. The publication series is primarily targeted at professionals. Economic Memo is published continuously.

WORKING PAPER

Working Paper presents research projects by economists in Danmarks Nationalbank and their associates. The series is primarily targeted at professionals and people with an interest for academia. Working Paper is published continuously.

DANMARKS NATIONALBANK
LANGELINIE ALLÉ 47
DK-2100 COPENHAGEN Ø
WWW.NATIONALBANKEN.DK

Danmarks Nationalbank's Economic Memos are published at www.nationalbanken.dk. A free electronic subscription is also available at the website. The subscriber receives an e-mail notification whenever a new Economic Memo is published.

Please direct any enquiries directly to the contributors or to Danmarks Nationalbank, Communications, *Kommunikation@nationalbanken.dk*.

DANMARKS NATIONALBANK