Nowcasting food inflation with a massive amount of online prices

Paweł Macias, Damian Stelmasiak & **Karol Szafranek**

Narodowy Bank Polski

The views expressed herein are those of their authors and not necessarily the views of Narodowy Bank Polski.

9th Summer Workshop of Narodowy Bank Polski
June 28-30, 2021
Virtual Meeting

Outline

1 Introduction

2 Literature review

3 Methodology

4 Results

5 Conclusions

Outline

1 Introduction

2 Literature review

3 Methodology

4 Results

5 Conclusions

Summary

1. Precise nowcasts are crucial for providing accurate forecasts (Faust and Wright, 2013).
2. Analysing online prices gaining a lot of momentum in recent years.
3. The research is spearheaded by studies within the Billion Prices Project, launched at MIT in 2008 (Cavallo & Ribogon, 2016).
4. We show that online prices can be very effective in nowcasting inflation and macroeconomic practitioners can make use of them already after a couple of months of data collection.

NBP Narodowy Bank Polski

Research question

- Can we draw from the Internet to successfully nowcast inflation?

Of course, we can.

- Is it worth it?

Entry cost is high, but ultimately macroeconomic practitioners should complement their analysis with online data.

Research question

- Can we draw from the Internet to successfully nowcast inflation?

Of course, we can.

- Is it worth it?

Entry cost is high, but ultimately macroeconomic practitioners should complement their analysis with online data.

Research question

- Can we draw from the Internet to successfully nowcast inflation?

Of course, we can.

- Is it worth it?

Entry cost is high, but ultimately macroeconomic practitioners should complement their analysis with online data.

NBP Narodowy Bank Polski

Research question

- Can we draw from the Internet to successfully nowcast inflation?

Of course, we can.

- Is it worth it?

Entry cost is high, but ultimately macroeconomic practitioners should complement their analysis with online data.

NBP Narodowy Bank Polski

This paper

1. We focus on inflation nowcasting.
2. We employ a unique, extensive dataset of online food and non-alcoholic beverages prices gathered from web since 2009.
3. Our database contains 159 millions of prices for around 640 thousands of products.
4. We perform a real-time nowcasting experiment among popular, simple univariate approaches using highly disaggregated framework.

This paper

1. We focus on inflation nowcasting.
2. We employ a unique, extensive dataset of online food and non-alcoholic beverages prices gathered from web since 2009.
3. Our database contains 159 millions of prices for around 640 thousands of products.
4. We perform a real-time nowcasting experiment among popular, simple univariate approaches using highly disaggregated framework.

NBP Narodowy Bank Polski

This paper

1. We focus on inflation nowcasting.
2. We employ a unique, extensive dataset of online food and non-alcoholic beverages prices gathered from web since 2009.
3. Our database contains 159 millions of prices for around 640 thousands of products.
4. We perform a real-time nowcasting experiment among popular, simple univariate approaches using highly disaggregated framework.

This paper

1. We focus on inflation nowcasting.
2. We employ a unique, extensive dataset of online food and non-alcoholic beverages prices gathered from web since 2009.
3. Our database contains 159 millions of prices for around 640 thousands of products.
4. We perform a real-time nowcasting experiment among popular, simple univariate approaches using highly disaggregated framework.

Why should we deal with online prices?

- Online prices are appealing for macroeconomic practitioners:

  1. delay in data collection is marginal – prices can be observed in real-time, downloaded remotely in any frequency.
  2. researcher can quickly obtain massive amount of data.
  3. there are limited costs of data retrieval in comparison to scanner data.
  4. additional information can complement the analysis (product description, availability, discounts) on an international micro level on price formation, price stickiness, price synchronization, market segmentation, shock transmission, tax changes, demand and supply disruption.

NBP Narodowy Bank Polski

Why should we deal with online prices?

- Online prices are appealing for macroeconomic practitioners:

  1. delay in data collection is marginal – prices can be observed in real-time, downloaded remotely in any frequency.
  2. researcher can quickly obtain massive amount of data.
  3. there are limited costs of data retrieval in comparison to scanner data.
  4. additional information can complement the analysis (product description, availability, discounts) on an international micro level on price formation, price stickiness, price synchronization, market segmentation, shock transmission, tax changes, demand and supply disruption.

NBP Narodowy Bank Polski

Why should we deal with online prices?

- Online prices are appealing for macroeconomic practitioners:
  1. delay in data collection is marginal – prices can be observed in real-time, downloaded remotely in any frequency.
  2. researcher can quickly obtain massive amount of data.
  3. there are limited costs of data retrieval in comparison to scanner data.
  4. additional information can complement the analysis (product description, availability, discounts) on an international micro level on price formation, price stickiness, price synchronization, market segmentation, shock transmission, tax changes, demand and supply disruption.

NBP Narodowy Bank Polski

Why should we deal with online prices?

- Online prices are appealing for macroeconomic practitioners:
  1. delay in data collection is marginal – prices can be observed in real-time, downloaded remotely in any frequency.
  2. researcher can quickly obtain massive amount of data.
  3. there are limited costs of data retrieval in comparison to scanner data.
  4. additional information can complement the analysis (product description, availability, discounts) on an international micro level on price formation, price stickiness, price synchronization, market segmentation, shock transmission, tax changes, demand and supply disruption.

NBP Narodowy Bank Polski

Why should we deal with online prices?

- Online prices are appealing for macroeconomic practitioners:
  1. delay in data collection is marginal – prices can be observed in real-time, downloaded remotely in any frequency.
  2. researcher can quickly obtain massive amount of data.
  3. there are limited costs of data retrieval in comparison to scanner data.
  4. additional information can complement the analysis (product description, availability, discounts) on an international micro level on price formation, price stickiness, price synchronization, market segmentation, shock transmission, tax changes, demand and supply disruption.

Our contribution

- In the paper we:
    1. discuss how macroeconomic practitioners can improve inflation nowcasts.
    2. provide evidence for a large number of highly disaggregated inflation components.
    3. study the usefulness of online prices for a small, emerging economy.
    4. report how forecasting errors are related to the variability in observed prices.
    5. provide guidelines for data curation and its importance in the nowcasting process.
    6. examine the accuracy of nowcasting with online prices during COVID-19 pandemic.

NBP Narodowy Bank Polski

Our contribution

- In the paper we:
  1. discuss how macroeconomic practitioners can improve inflation nowcasts.
  2. provide evidence for a large number of highly disaggregated inflation components.
  3. study the usefulness of online prices for a small, emerging economy.
  4. report how forecasting errors are related to the variability in observed prices.
  5. provide guidelines for data curation and its importance in the nowcasting process.
  6. examine the accuracy of nowcasting with online prices during COVID-19 pandemic.

NBP Narodowy Bank Polski

Our contribution

- In the paper we:
  1. discuss how macroeconomic practitioners can improve inflation nowcasts.
  2. provide evidence for a large number of highly disaggregated inflation components.
  3. study the usefulness of online prices for a small, emerging economy.
  4. report how forecasting errors are related to the variability in observed prices.
  5. provide guidelines for data curation and its importance in the nowcasting process.
  6. examine the accuracy of nowcasting with online prices during COVID-19 pandemic.

NBP Narodowy Bank Polski

Our contribution

- In the paper we:
  1. discuss how macroeconomic practitioners can improve inflation nowcasts.
  2. provide evidence for a large number of highly disaggregated inflation components.
  3. study the usefulness of online prices for a small, emerging economy.
  4. report how forecasting errors are related to the variability in observed prices.
  5. provide guidelines for data curation and its importance in the nowcasting process.
  6. examine the accuracy of nowcasting with online prices during COVID-19 pandemic.

NBP Narodowy Bank Polski

Our contribution

- In the paper we:
  1. discuss how macroeconomic practitioners can improve inflation nowcasts.
  2. provide evidence for a large number of highly disaggregated inflation components.
  3. study the usefulness of online prices for a small, emerging economy.
  4. report how forecasting errors are related to the variability in observed prices.
  5. provide guidelines for data curation and its importance in the nowcasting process.
  6. examine the accuracy of nowcasting with online prices during COVID-19 pandemic.

NBP Narodowy Bank Polski

Our contribution

- In the paper we:
  1. discuss how macroeconomic practitioners can improve inflation nowcasts.
  2. provide evidence for a large number of highly disaggregated inflation components.
  3. study the usefulness of online prices for a small, emerging economy.
  4. report how forecasting errors are related to the variability in observed prices.
  5. provide guidelines for data curation and its importance in the nowcasting process.
  6. examine the accuracy of nowcasting with online prices during COVID-19 pandemic.

NBP Narodowy Bank Polski

Key results

- We show that:
  1. pure estimates of online price changes is already effective in nowcasting food inflation.
  2. incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
  3. this approach outperforms a variety of frameworks, including judgemental methods.
  4. marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
  5. during COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
  6. meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

Key results

- We show that:

1. pure estimates of online price changes is already effective in nowcasting food inflation.
2. incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
3. this approach outperforms a variety of frameworks, including judgemental methods.
4. marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
5. during COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
6. meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

Key results

- We show that:
    1. pure estimates of online price changes is already effective in nowcasting food inflation.
    2. incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
    3. this approach outperforms a variety of frameworks, including judgemental methods.
    4. marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
    5. during COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
    6. meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

NBP Narodowy Bank Polski

Key results

- We show that:
  1. pure estimates of online price changes is already effective in nowcasting food inflation.
  2. incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
  3. this approach outperforms a variety of frameworks, including judgemental methods.
  4. marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
  5. during COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
  6. meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

NBP Narodowy Bank Polski

Key results

- We show that:
  1. pure estimates of online price changes is already effective in nowcasting food inflation.
  2. incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
  3. this approach outperforms a variety of frameworks, including judgemental methods.
  4. marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
  5. during COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
  6. meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

NBP Narodowy Bank Polski

Key results

- We show that:
  1. pure estimates of online price changes is already effective in nowcasting food inflation.
  2. incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
  3. this approach outperforms a variety of frameworks, including judgemental methods.
  4. marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
  5. during COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
  6. meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

Outline

1 Introduction

2 Literature review

3 Methodology

4 Results

5 Conclusions

Forecasting inflation

- Forecasting inflation is a daunting challenge.
- Naive benchmarks extrapolating past inflation trends or accounting for the slowly varying local mean are hard to beat (Atkeson & Ohanian, 2001; Stock & Watson, 2007; Faust and Wright, 2013).
- Judgemental forecasts often provide superior accuracy (Faust & Wright, 2013).
- A plethora of sophisticated models has been developed to forecast inflation and a just review of this strand of literature seems infeasible.
- Our understanding of this research is that inflation nowcasts are often overlooked.
- Yet, the success of preparing accurate inflation forecast heavily relies on accurately pinning down the inflation nowcast (Faust & Wright, 2013).
- Sophisticated models have been developed for nowcasting inflation (Monteforte and Moretti, 2013; Modugno, 2013; Knotek & Zaman, 2017; Tallman and Zaman, 2020).

NBP Narodowy Bank Polski

Forecasting inflation

- Forecasting inflation is a daunting challenge.
- Naive benchmarks extrapolating past inflation trends or accounting for the slowly varying local mean are hard to beat (Atkeson & Ohanian, 2001; Stock & Watson, 2007; Faust and Wright, 2013).
- Judgemental forecasts often provide superior accuracy (Faust & Wright, 2013).
- A plethora of sophisticated models has been developed to forecast inflation and a just review of this strand of literature seems infeasible.
- Our understanding of this research is that inflation nowcasts are often overlooked.
- Yet, the success of preparing accurate inflation forecast heavily relies on accurately pinning down the inflation nowcast (Faust & Wright, 2013).
- Sophisticated models have been developed for nowcasting inflation (Monteforte and Moretti, 2013; Modugno, 2013; Knotek & Zaman, 2017; Tallman and Zaman, 2020).

Forecasting inflation

- Forecasting inflation is a daunting challenge.
- Naive benchmarks extrapolating past inflation trends or accounting for the slowly varying local mean are hard to beat (Atkeson & Ohanian, 2001; Stock & Watson, 2007; Faust and Wright, 2013).
- Judgemental forecasts often provide superior accuracy (Faust & Wright, 2013).
- A plethora of sophisticated models has been developed to forecast inflation and a just review of this strand of literature seems infeasible.
- Our understanding of this research is that inflation nowcasts are often overlooked.
- Yet, the success of preparing accurate inflation forecast heavily relies on accurately pinning down the inflation nowcast (Faust & Wright, 2013).
- Sophisticated models have been developed for nowcasting inflation (Monteforte and Moretti, 2013; Modugno, 2013; Knotek & Zaman, 2017; Tallman and Zaman, 2020).

NBP Narodowy Bank Polski

Forecasting inflation

- Forecasting inflation is a daunting challenge.
- Naive benchmarks extrapolating past inflation trends or accounting for the slowly varying local mean are hard to beat (Atkeson & Ohanian, 2001; Stock & Watson, 2007; Faust and Wright, 2013).
- Judgemental forecasts often provide superior accuracy (Faust & Wright, 2013).
- A plethora of sophisticated models has been developed to forecast inflation and a just review of this strand of literature seems infeasible.
- Our understanding of this research is that inflation nowcasts are often overlooked.
- Yet, the success of preparing accurate inflation forecast heavily relies on accurately pinning down the inflation nowcast (Faust & Wright, 2013).
- Sophisticated models have been developed for nowcasting inflation (Monteforte and Moretti, 2013; Modugno, 2013; Knotek & Zaman, 2017; Tallman and Zaman, 2020).

NBP Narodowy Bank Polski

Forecasting inflation

- Forecasting inflation is a daunting challenge.
- Naive benchmarks extrapolating past inflation trends or accounting for the slowly varying local mean are hard to beat (Atkeson & Ohanian, 2001; Stock & Watson, 2007; Faust and Wright, 2013).
- Judgemental forecasts often provide superior accuracy (Faust & Wright, 2013).
- A plethora of sophisticated models has been developed to forecast inflation and a just review of this strand of literature seems infeasible.
- Our understanding of this research is that inflation nowcasts are often overlooked.
- Yet, the success of preparing accurate inflation forecast heavily relies on accurately pinning down the inflation nowcast (Faust & Wright, 2013).
- Sophisticated models have been developed for nowcasting inflation (Monteforte and Moretti, 2013; Modugno, 2013; Knotek & Zaman, 2017; Tallman and Zaman, 2020).

NBP Narodowy Bank Polski

Forecasting inflation

- Forecasting inflation is a daunting challenge.
- Naive benchmarks extrapolating past inflation trends or accounting for the slowly varying local mean are hard to beat (Atkeson & Ohanian, 2001; Stock & Watson, 2007; Faust and Wright, 2013).
- Judgemental forecasts often provide superior accuracy (Faust & Wright, 2013).
- A plethora of sophisticated models has been developed to forecast inflation and a just review of this strand of literature seems infeasible.
- Our understanding of this research is that inflation nowcasts are often overlooked.
- Yet, the success of preparing accurate inflation forecast heavily relies on accurately pinning down the inflation nowcast (Faust & Wright, 2013).
- Sophisticated models have been developed for nowcasting inflation (Monteforte and Moretti, 2013; Modugno, 2013; Knotek & Zaman, 2017; Tallman and Zaman, 2020).

NBP Narodowy Bank Polski

Forecasting inflation

- Forecasting inflation is a daunting challenge.
- Naive benchmarks extrapolating past inflation trends or accounting for the slowly varying local mean are hard to beat (Atkeson & Ohanian, 2001; Stock & Watson, 2007; Faust and Wright, 2013).
- Judgemental forecasts often provide superior accuracy (Faust & Wright, 2013).
- A plethora of sophisticated models has been developed to forecast inflation and a just review of this strand of literature seems infeasible.
- Our understanding of this research is that inflation nowcasts are often overlooked.
- Yet, the success of preparing accurate inflation forecast heavily relies on accurately pinning down the inflation nowcast (Faust & Wright, 2013).
- Sophisticated models have been developed for nowcasting inflation (Monteforte and Moretti, 2013; Modugno, 2013; Knotek & Zaman, 2017; Tallman and Zaman, 2020).

NBP Narodowy Bank Polski

Usefulness of online data

- Literature on online prices is mushrooming (Lunnemann and Wintr, 2011; Cavallo, 2013; Cavallo & Ribogon, 2016; Cavallo, 2017; Gorodnichenko & Talavera, 2017; Cavallo, 2018; Gorodnichenko et al., 2018).
- Online prices are increasingly included in the compilation of the CPI (in the US, the UK, the Netherlands, New Zealand and Norway).
- Evidence on the usefulness of scraped data in forecasting inflation remains scarce (Aparicio & Bertolotto, 2020).
  - Scraped data from July 2008 to September 2016.
  - Data for 10 advanced economies.
  - Parsimonious models with online prices beat traditional benchmarks and two leading survey of professional forecasters.
- Our work is conceptually similar but we:
  - provide evidence for a large number of highly disaggregated components,
  - study the usefulness of online prices for an emerging economy,
  - show how forecasting errors are related to the variability of online prices,
  - discuss the importance of data curation,
  - examine the nowcast accuracy during COVID-19.

NBP Narodowy Bank Polski

Usefulness of online data

- Literature on online prices is mushrooming (Lunnemann and Wintr, 2011; Cavallo, 2013; Cavallo & Ribogon, 2016; Cavallo, 2017; Gorodnichenko & Talavera, 2017; Cavallo, 2018; Gorodnichenko et al., 2018).
- Online prices are increasingly included in the compilation of the CPI (in the US, the UK, the Netherlands, New Zealand and Norway).
- Evidence on the usefulness of scraped data in forecasting inflation remains scarce (Aparicio & Bertolotto, 2020).
  - Scraped data from July 2008 to September 2016.
  - Data for 10 advanced economies.
  - Parsimonious models with online prices beat traditional benchmarks and two leading survey of professional forecasters.
- Our work is conceptually similar but we:
  - provide evidence for a large number of highly disaggregated components,
  - study the usefulness of online prices for an emerging economy,
  - show how forecasting errors are related to the variability of online prices,
  - discuss the importance of data curation,
  - examine the nowcast accuracy during COVID-19.

Usefulness of online data

- Literature on online prices is mushrooming (Lunnemann and Wintr, 2011; Cavallo, 2013; Cavallo & Ribogon, 2016; Cavallo, 2017; Gorodnichenko & Talavera, 2017; Cavallo, 2018; Gorodnichenko et al., 2018).
- Online prices are increasingly included in the compilation of the CPI (in the US, the UK, the Netherlands, New Zealand and Norway).
- Evidence on the usefulness of scraped data in forecasting inflation remains scarce (Aparicio & Bertolotto, 2020).
  - Scraped data from July 2008 to September 2016.
  - Data for 10 advanced economies.
  - Parsimonious models with online prices beat traditional benchmarks and two leading survey of professional forecasters.
- Our work is conceptually similar but we:
  - provide evidence for a large number of highly disaggregated components,
  - study the usefulness of online prices for an emerging economy,
  - show how forecasting errors are related to the variability of online prices,
  - discuss the importance of data curation,
  - examine the nowcast accuracy during COVID-19.

NBP Narodowy Bank Polski

Usefulness of online data

- Literature on online prices is mushrooming (Lunnemann and Wintr, 2011; Cavallo, 2013; Cavallo & Ribogon, 2016; Cavallo, 2017; Gorodnichenko & Talavera, 2017; Cavallo, 2018; Gorodnichenko et al., 2018).
- Online prices are increasingly included in the compilation of the CPI (in the US, the UK, the Netherlands, New Zealand and Norway).
- Evidence on the usefulness of scraped data in forecasting inflation remains scarce (Aparicio & Bertolotto, 2020).
  - Scraped data from July 2008 to September 2016.
  - Data for 10 advanced economies.
  - Parsimonious models with online prices beat traditional benchmarks and two leading survey of professional forecasters.
- Our work is conceptually similar but we:
  - provide evidence for a large number of highly disaggregated components,
  - study the usefulness of online prices for an emerging economy,
  - show how forecasting errors are related to the variability of online prices,
  - discuss the importance of data curation,
  - examine the nowcast accuracy during COVID-19.

NBP Narodowy Bank Polski

E-commerce market in Poland

1. Polish e-commerce market is relatively young, still in the development stage.
2. It is characterized by an impetuous growth in 2004-2018, with CAGR at 22.5%.
3. Online retail sales skyrocketed temporarily during COVID-19 lockdown.
4. The percentage of individuals buying online rises systematically.
5. The revenue in the e-commerce market is projected to reach around $13.5 billion in 2021 and increase by 7.3% until 2025.

E-commerce market in Poland

1. Polish e-commerce market is relatively young, still in the development stage.
2. It is characterized by an impetuous growth in 2004-2018, with CAGR at 22.5%.
3. Online retail sales skyrocketed temporarily during COVID-19 lockdown.
4. The percentage of individuals buying online rises systematically.
5. The revenue in the e-commerce market is projected to reach around $13.5 billion in 2021 and increase by 7.3% until 2025.

E-commerce market in Poland

1. Polish e-commerce market is relatively young, still in the development stage.
2. It is characterized by an impetuous growth in 2004-2018, with CAGR at 22.5%.
3. Online retail sales skyrocketed temporarily during COVID-19 lockdown.
4. The percentage of individuals buying online rises systematically.
5. The revenue in the e-commerce market is projected to reach around $13.5 billion in 2021 and increase by 7.3% until 2025.

E-commerce market in Poland

1. Polish e-commerce market is relatively young, still in the development stage.
2. It is characterized by an impetuous growth in 2004-2018, with CAGR at 22.5%.
3. Online retail sales skyrocketed temporarily during COVID-19 lockdown.
4. The percentage of individuals buying online rises systematically.
5. The revenue in the e-commerce market is projected to reach around $13.5 billion in 2021 and increase by 7.3% until 2025.

E-commerce market in Poland

1. Polish e-commerce market is relatively young, still in the development stage.
2. It is characterized by an impetuous growth in 2004-2018, with CAGR at 22.5%.
3. Online retail sales skyrocketed temporarily during COVID-19 lockdown.
4. The percentage of individuals buying online rises systematically.
5. The revenue in the e-commerce market is projected to reach around $13.5 billion in 2021 and increase by 7.3% until 2025.

Outline

1 Introduction

2 Literature review

3 Methodology

4 Results

5 Conclusions

Online data

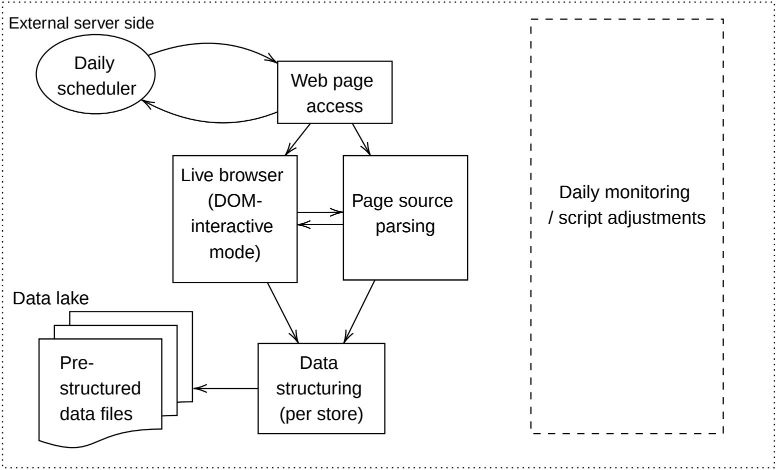
- E-CPI project aimed at collecting weekly online prices launched in December, 2009.
- Frequency of data collection changed to daily since 2017.
- Systematically expanding product categories to account for clothing, footwear, products, electronics, drugs, air plain tickets.

Store ID	q	q_s	n_p	$\bar{n}_p$	$\bar{n}_d$	$med\{n_d\}$
1	82,995	19,250	13,394,437	7,327	161	42
2	56,739	16,368	14,721,627	8,544	260	76
3	76,652	13,007	25,600,013	17,534	334	222
4	50,954	15,935	17,196,321	11,495	341	164
5	132,543	26,175	43,387,750	23,709	327	187
6	21,658	7,381	6,346,966	9,544	302	266
7	103,234	34,054	16,708,475	9,170	175	87
8	119,168	27,544	21,630,105	13,213	182	34
Total	643,943	159,714	158,985,694	12,759	250	102

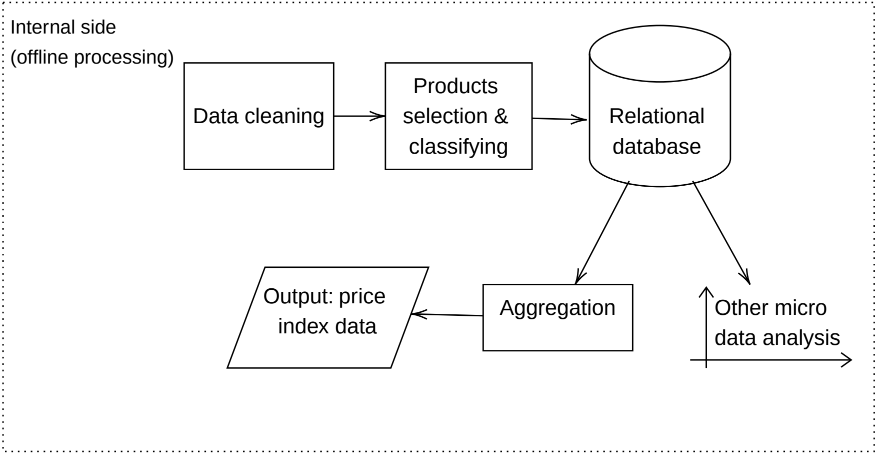
q – number of products, q_s – number of selected products, n_p – number of observed prices, $\bar{n}_p$ – average number of prices per day, $\bar{n}_d$ – mean number of days price is observed, $med\{n_d\}$ – median number of days price is observed.

NBP Narodowy Bank Polski

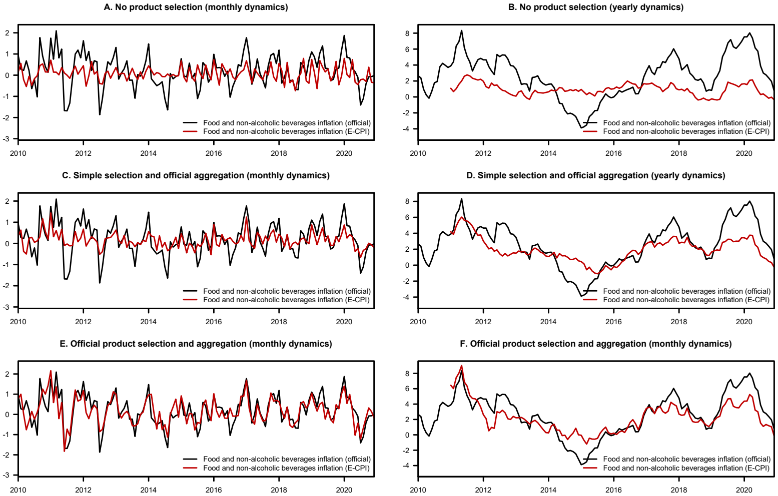
Web scraping procedure – external server side



Web scraping procedure – offline processing



In-sample tracking accuracy of online data



Nowcasting competition (1)

- Inflation rate as the monthly, non-seasonally adjusted change in prices as the variable of interest.
- Change in the online price index defined as $o_{c,t} = 100 * O_{c,t}/O_{c,t-1}$ as the auxiliary independent variable.
- Baseline approach – recursive estimation strategy making use of expanding window.
  - Estimation sample spanning the period January 1999 - December 2016.
  - Evaluation sample spanning the period January 2017 - December 2020.
- Nowcast evaluation with MFE, RMSFE and Diebold-Mariano tests.
  - MFE reported in levels.
  - RMSFE reported as ratios - a value above 1 (below 1) that the competing approach produces on average less (more) accurate nowcasts than EC^{SX}.
- Sensitivity analysis for:
  - rolling window estimation,
  - simplistic data curation,
  - forecast combinations,
  - the COVID-19 period (i.e. January 2020 - December 2020).

NBP Narodowy Bank Polski

Nowcasting competition (1)

- Inflation rate as the monthly, non-seasonally adjusted change in prices as the variable of interest.
- Change in the online price index defined as $o_{c,t} = 100 * O_{c,t}/O_{c,t-1}$ as the auxiliary independent variable.
- Baseline approach – recursive estimation strategy making use of expanding window.
  - Estimation sample spanning the period January 1999 - December 2016.
  - Evaluation sample spanning the period January 2017 - December 2020.
- Nowcast evaluation with MFE, RMSFE and Diebold-Mariano tests.
  - MFE reported in levels.
  - RMSFE reported as ratios - a value above 1 (below 1) that the competing approach produces on average less (more) accurate nowcasts than EC^{SX}.
- Sensitivity analysis for:
  - rolling window estimation,
  - simplistic data curation,
  - forecast combinations,
  - the COVID-19 period (i.e. January 2020 - December 2020).

NBP Narodowy Bank Polski

Nowcasting competition (1)

- Inflation rate as the monthly, non-seasonally adjusted change in prices as the variable of interest.
- Change in the online price index defined as $o_{c,t} = 100 * O_{c,t}/O_{c,t-1}$ as the auxiliary independent variable.
- Baseline approach – recursive estimation strategy making use of expanding window.
  - Estimation sample spanning the period January 1999 - December 2016.
  - Evaluation sample spanning the period January 2017 - December 2020.
- Nowcast evaluation with MFE, RMSFE and Diebold-Mariano tests.
  - MFE reported in levels.
  - RMSFE reported as ratios - a value above 1 (below 1) that the competing approach produces on average less (more) accurate nowcasts than EC^{SX}.
- Sensitivity analysis for:
  - rolling window estimation,
  - simplistic data curation,
  - forecast combinations,
  - the COVID-19 period (i.e. January 2020 - December 2020).

NBP Narodowy Bank Polski

Nowcasting competition (1)

- Inflation rate as the monthly, non-seasonally adjusted change in prices as the variable of interest.
- Change in the online price index defined as $o_{c,t} = 100 * O_{c,t}/O_{c,t-1}$ as the auxiliary independent variable.
- Baseline approach – recursive estimation strategy making use of expanding window.
  - Estimation sample spanning the period January 1999 - December 2016.
  - Evaluation sample spanning the period January 2017 - December 2020.
- Nowcast evaluation with MFE, RMSFE and Diebold-Mariano tests.
  - MFE reported in levels.
  - RMSFE reported as ratios - a value above 1 (below 1) that the competing approach produces on average less (more) accurate nowcasts than EC^{SX}.
- Sensitivity analysis for:
  - rolling window estimation,
  - simplistic data curation,
  - forecast combinations,
  - the COVID-19 period (i.e. January 2020 - December 2020).

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i} \varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i} \varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0, \sigma_c^2)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^*_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B} \beta_{c,i} o_{c,t-i} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i} \varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i} \varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0, \sigma_c^2)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0, \sigma_c^2)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^*_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B} \beta_{c,i} o_{c,t-i} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0, \sigma_c^2)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

NBP Narodowy Bank Polski

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0, \sigma^2_c)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^*_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B} \beta_{c,i} o_{c,t-i} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0, \sigma^2_c)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

NBP Narodowy Bank Polski

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0, \sigma_c^2)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^{*}_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B} \beta_{c,i} o_{c,t-i} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0, \sigma_c^2)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i} \varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i} \varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0, \sigma_c^2)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^{*}_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B} \beta_{c,i} o_{c,t-i} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i} \varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i} \varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0, \sigma_c^2)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

NBP Narodowy Bank Polski

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0, \sigma^2_c)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^*_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B} \beta_{c,i} o_{c,t-i} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0, \sigma^2_c)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

NBP Narodowy Bank Polski

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0, \sigma^2_c)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^*_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B} \beta_{c,i} o_{c,t-i} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0, \sigma^2_c)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

NBP Narodowy Bank Polski

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p}\phi_{c,i}p_{c,t-i} + \sum_{i=1}^{q}\theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P}\Phi_{c,i}p_{c,t-i*12} + \sum_{i=1}^{Q}\Theta_{c,i}\varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0,\sigma^2_c)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^{*}_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B}\beta_{c,i}o_{c,t-i} + \sum_{i=1}^{p}\phi_{c,i}p_{c,t-i} + \sum_{i=1}^{q}\theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P}\Phi_{c,i}p_{c,t-i*12} + \sum_{i=1}^{Q}\Theta_{c,i}\varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0,\sigma^2_c)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

NBP Narodowy Bank Polski

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0, \sigma^2_c)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^*_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B} \beta_{c,i} o_{c,t-i} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0, \sigma^2_c)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0,\sigma_c^2)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^{*}_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B} \beta_{c,i} o_{c,t-i} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0,\sigma_c^2)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

NBP Narodowy Bank Polski

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0, \sigma^2_c)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^{*}_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B} \beta_{c,i} o_{c,t-i} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0, \sigma^2_c)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

NBP Narodowy Bank Polski

Nowcasting competition (2)

- Model entering the nowcasting competition:
  1. The simple random walk: $p_{c,t+1} = p_{c,t}$.
  2. The random walk à la Atkeson & Ohanian (2001): $p_{c,t+1} = \frac{1}{12}\sum_{j=1}^{12} p^{SA}_{c,t-j+1} + \hat{s}^{TS}_{c,t+1}$.
  3. The best SARMA model based on in-sample fit (BS^{IS}): $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$, assuming that $\varepsilon_{c,t} \sim NIID(0,\sigma_c^2)$ chosen from 64 specifications using BIC.
  4. The best SARMA model based on out-of-sample accuracy (BS^{OS}) using the minimal RMSFE criterion calculated on a pseudo validation set.
  5. The pure, ex-post E-CPI nowcast: $p_{c,t+1} = o_{c,t+1}$.
  6. The pure, real-time E-CPI nowcast (EC^{RT}): $p_{c,t+1} = o^{*}_{c,t+1}$.
  7. The best SARMAX model with online prices as the exogenous variable (EC^{SX}). The specification is the following: $p_{c,t} = \mu_c + \varepsilon_{c,t} + \sum_{i=0}^{B} \beta_{c,i} o_{c,t-i} + \sum_{i=1}^{p} \phi_{c,i} p_{c,t-i} + \sum_{i=1}^{q} \theta_{c,i}\varepsilon_{c,t-i} + \sum_{i=1}^{P} \Phi_{c,i} p_{c,t-i*12} + \sum_{i=1}^{Q} \Theta_{c,i}\varepsilon_{c,t-i*12}$ and $\varepsilon_{c,t} \sim NIID(0,\sigma_c^2)$ chosen from 256 specifications using the minimal RMSFE criterion calculated on a pseudo validation set.
  8. The simple combination of the EC^{SX} models (EC^{MC}).
  9. The combination of BS^{OS} models with equal weights (BS^{MC}).
  10. The combination of EC^{SX} models with weight inversely proportional to RMSFE (EC^{MCI}).
  11. The judgemental forecast (JD).

NBP Narodowy Bank Polski

Outline

1 Introduction

2 Literature review

3 Methodology

4 Results

5 Conclusions

Baseline results – the aggregate, recursive estimation

Error	Online prices			Traditional benchmarks				Forecast combinations		
	EC^{SX}	EC^{EX}	EC^{RT}	RW	AO^{SA}	BS^{IS}	BS^{OS}	BS^{MC}	EC^{MC}	EC^{MCI}
					Recursive estimation					
MFE	0.018	-0.106	-0.093	0.042	0.118	0.109	0.111	0.134	0.025	0.022
RMSFE	0.338	1.071	1.082	1.784^{a}	1.528^{a}	1.446^{a}	1.524^{a}	1.621^{a}	1.181^{b}	1.176^{b}
					Rolling estimation					
MFE	0.019	-0.106	-0.093	0.042	0.118	0.041	0.017	0.025	0.080	0.019
RMSFE	0.427	0.848	0.857	1.413^{b}	1.209^{c}	1.362^{b}	1.381^{b}	1.324^{b}	1.003	0.965
					COVID-19					
MFE	0.146	-0.076	-0.078	0.084	0.362	0.401	0.394	0.394	0.212	0.214
RMSFE	0.340	1.312	1.332	1.567	1.575	1.727	1.902	2.048	1.411	1.398

MFE reported in levels. RMSFE reported as ratios - a value above 1 (below 1) that the competing approach produces on average less (more) accurate nowcasts than EC^{SX}. ^{a} denotes significance at the 1 percent level, ^{b} denotes significance at the 5 percent level, ^{c} denotes significance at the 10 percent level.

NBP Narodowy Bank Polski

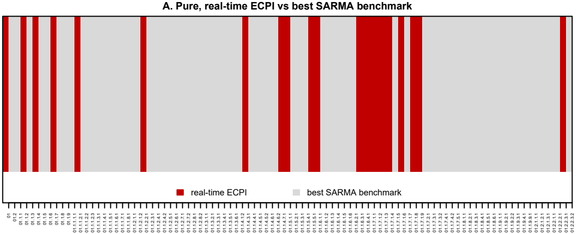
Baseline results – medium level disaggregation

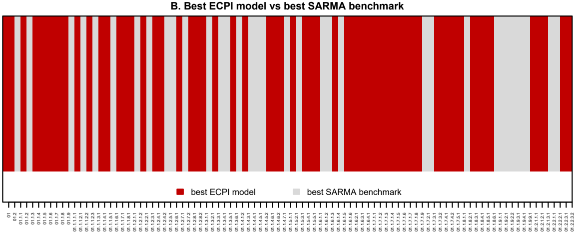
COICOP	Online prices			Traditional benchmarks				Forecast combinations		
	EC^{SX}	EC^{EX}	EC^{RT}	RW	AO^{SA}	BS^{IS}	BS^{OS}	BS^{MC}	EC^{MC}	EC^{MCI}
01.	0.338	1.071	1.082	1.784[a]	1.528[a]	1.446[a]	1.524[a]	1.621[a]	1.181[b]	1.176[b]
01.1.1	0.264	1.295[b]	1.338[b]	1.189[b]	1.153[c]	0.987	0.991	0.975	0.955	0.949
01.1.2	0.747	0.825	0.952	1.483[a]	1.285[b]	1.153[a]	1.151[b]	1.166[a]	1.062	1.061
01.1.3	0.319	2.070[a]	2.242[a]	1.621[a]	0.983	1.093	0.991	1.080	1.017	1.019
01.1.4	0.536	0.892	0.840	1.630	1.489[b]	1.400	1.349	1.411[c]	1.065	1.063
01.1.5	1.237	0.818[b]	1.072	1.400[a]	1.155	1.083[b]	1.030	1.121[b]	1.071[c]	1.068[c]
01.1.6	2.085	1.126	1.285	1.460[a]	1.224[b]	1.103[c]	1.136[c]	1.227[b]	1.088	1.070
01.1.7	1.737	1.011	0.976	2.432[a]	1.707[a]	1.655[a]	1.686[a]	1.734[a]	0.957	0.954
01.1.8	0.421	1.272[a]	1.536[a]	1.573[a]	1.235[b]	1.050[b]	1.061	1.115	1.002	1.000
01.1.9	0.260	1.474[b]	1.783[b]	2.130[a]	1.101	1.088[b]	1.122	1.222[b]	1.066	1.061
01.2	0.234	1.845[a]	1.743[a]	1.940[a]	1.103	1.080	1.104	1.205[b]	1.048	1.046

MFE reported in levels. RMSFE reported as ratios - a value above 1 (below 1) that the competing approach produces on average less (more) accurate nowcasts than EC^{SX}. [a] denotes significance at the 1 percent level, [b] denotes significance at the 5 percent level, [c] denotes significance at the 10 percent level. 01. denotes *Food and non-alcoholic beverages*, 01.1.1 denotes *Bread and cereals*, 01.1.2 denotes *Meat*, 01.1.3 denotes *Fish and seafood*, 01.1.4 denotes *Milk, cheese and eggs*, 01.1.5 denotes *Oils and fats*, 01.1.6 denotes *Fruits*, 01.1.7 denotes *Vegetables*, 01.1.8 denotes *Sugar, jam, honey, chocolate and confectionery*, 01.1.9 denotes *Food products, n.e.c.* and 01.2 denotes *Non-alcoholic beverages*.

NBP Narodowy Bank Polski

Baseline results – low level disaggregation





The impact of price dispersion on the relative nowcast accuracy

	$RMSFE_R$
log(SD)	-0.065^a
01.1.1	0.111^b
01.1.2	0.157^a
01.1.3	0.182^a
01.1.4	0.099^b
01.1.5	0.133^a
01.1.6	0.189^a
01.1.7	0.108^b
01.1.8	0.117^a
01.1.9	0.157^a
01.2	0.089^b
Constant	-0.157^a
Observations	95
R^2	0.368
F Statistic	4.390^a (df = 11; 83)

01.1.1 denotes *Bread and cereals*, 01.1.2 denotes *Meat*, 01.1.3 denotes *Fish and seafood*, 01.1.4 denotes *Milk, cheese and eggs*, 01.1.5 denotes *Oils and fats*, 01.1.6 denotes *Fruits*, 01.1.7 denotes *Vegetables*, 01.1.8 denotes *Sugar, jam, honey, chocolate and confectionery*, 01.1.9 denotes *Food products, n.e.c.* and 01.2 denotes *Non-alcoholic beverages*. [a] denotes significance at the 1 percent level, [b] denotes significance at the 5 percent level, [c] denotes significance at the 10 percent level.

NBP Narodowy Bank Polski

Sensitivity analysis – the aggregate, rolling estimation

Error	Online prices			Traditional benchmarks				Forecast combinations		
	EC^{SX}	EC^{EX}	EC^{RT}	RW	AO^{SA}	BS^{IS}	BS^{OS}	BS^{MC}	EC^{MC}	EC^{MCI}
Recursive estimation										
MFE	0.018	-0.106	-0.093	0.042	0.118	0.109	0.111	0.134	0.025	0.022
RMSFE	0.338	1.071	1.082	1.784^a	1.528^a	1.446^a	1.524^a	1.621^a	1.181^b	1.176^b
Rolling estimation										
MFE	0.019	-0.106	-0.093	0.042	0.118	0.041	0.017	0.025	0.080	0.019
RMSFE	0.427	0.848	0.857	1.413^b	1.209^c	1.362^b	1.381^b	1.324^b	1.003	0.965
COVID-19										
MFE	0.146	-0.076	-0.078	0.084	0.362	0.401	0.394	0.394	0.212	0.214
RMSFE	0.340	1.312	1.332	1.567	1.575	1.727	1.902	2.048	1.411	1.398

MFE reported in levels. RMSFE reported as ratios - a value above 1 (below 1) that the competing approach produces on average less (more) accurate nowcasts than EC^{SX}. ^{a} denotes significance at the 1 percent level, ^{b} denotes significance at the 5 percent level, ^{c} denotes significance at the 10 percent level.

NBP Narodowy Bank Polski

Recursive versus rolling estimation

		EC^{SX}	EC^{MC}	EC^{MCI}	BS^{IS}	BS^{OS}	BS^{MC}	BS^{MCI}
RMSFE	rolling window	0.427	0.428	0.412	0.581	0.589	0.565	0.566
	expanding window	0.338	0.399	0.397	0.489	0.515	0.548	0.545
	ratio	26%	7%	4%	19%	14%	3%	4%
MFE	rolling	0.019	0.080	0.019	0.041	0.017	0.025	0.027
	expanding	0.018	0.025	0.022	0.109	0.111	0.134	0.131
	difference	0.001	0.054	-0.003	-0.068	-0.093	-0.109	-0.104

For the RMSFE statistics the ratio denotes the percent change in the RMSFE when the expanding windows estimation is switched to the rolling windows. For the MFE a simple difference is reported.

NBP Narodowy Bank Polski

The comparison between online data approach and judgemental nowcasts

	Recursive estimation (whole evaluation period)		Rolling estimation (whole evaluation period)		Recursive estimation COVID-19 period	
	EC^{SX}	JD	EC^{SX}	JD	EC^{SX}	JD
MFE	0.018	0.021	0.019	0.021	0.146	0.147
RMSFE	0.338	1.144[c]	0.427	0.905	0.340	0.965

MFE reported in levels. RMSFE reported as ratios - a value above 1 (below 1) that the competing approach produces on average less (more) accurate nowcasts than EC^{SX}. [a] denotes significance at the 1 percent level, [b] denotes significance at the 5 percent level, [c] denotes significance at the 10 percent level.

The importance of data curation

		RMSFE	MFE
Frameworks with online prices	EC^{SX}	0.338	0.018
	EC^{EX}	1.071	-0.106
	EC^{RT}	1.082	-0.093
Simple selection	EC^{SX}	1.299^{b}	0.036
	EC^{RT}	1.400^{a}	-0.117
Traditional benchmarks	RW	1.784^{a}	0.042
	AO^{SA}	1.528^{a}	0.118
	BS^{IS}	1.446^{a}	0.109
	BS^{OS}	1.524^{a}	0.111
Forecast combinations	BS^{MC}	1.621^{a}	0.134
	EC^{MC}	1.181^{b}	0.025
	EC^{MCI}	1.176^{b}	0.022

MFE reported in levels. RMSFE reported as ratios - a value above 1 (below 1) that the competing approach produces on average less (more) accurate nowcasts than EC^{SX}. Simple selection refers to the robustness check, where products are classified into respective groups using unsupervised learning based on word stems. ^{a} denotes significance at the 1 percent level, ^{b} denotes significance at the 5 percent level, ^{c} denotes significance at the 10 percent level.

NBP Narodowy Bank Polski

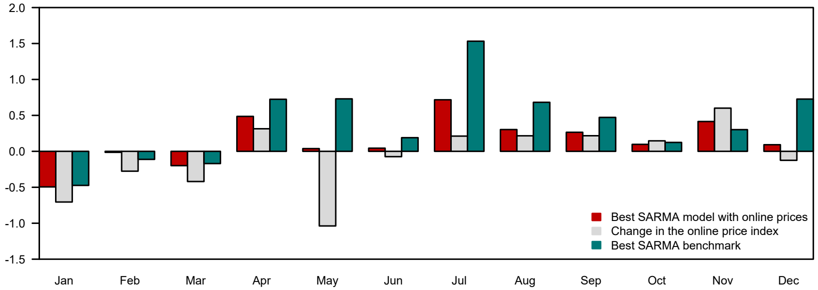
Sensitivity analysis – the aggregate, COVID-19 period (2020M1-2020M12)

Error	Online prices			Traditional benchmarks				Forecast combinations		
	EC^{SX}	EC^{EX}	EC^{RT}	RW	AO^{SA}	BS^{IS}	BS^{OS}	BS^{MC}	EC^{MC}	EC^{MCI}
				Recursive estimation						
MFE	0.018	-0.106	-0.093	0.042	0.118	0.109	0.111	0.134	0.025	0.022
RMSFE	0.338	1.071	1.082	1.784^{a}	1.528^{a}	1.446^{a}	1.524^{a}	1.621^{a}	1.181^{b}	1.176^{b}
				Rolling estimation						
MFE	0.019	-0.106	-0.093	0.042	0.118	0.041	0.017	0.025	0.080	0.019
RMSFE	0.427	0.848	0.857	1.413^{b}	1.209^{c}	1.362^{b}	1.381^{b}	1.324^{b}	1.003	0.965
				COVID-19						
MFE	0.146	-0.076	-0.078	0.084	0.362	0.401	0.394	0.394	0.212	0.214
RMSFE	0.340	1.312	1.332	1.567	1.575	1.727	1.902	2.048	1.411	1.398

MFE reported in levels. RMSFE reported as ratios - a value above 1 (below 1) that the competing approach produces on average less (more) accurate nowcasts than EC^{SX}. ^{a} denotes significance at the 1 percent level, ^{b} denotes significance at the 5 percent level, ^{c} denotes significance at the 10 percent level.

NBP Narodowy Bank Polski

Nowcasting performance during COVID-19



Nowcast errors in the subsequent months of 2020 for the EC^{SX} model, the EC^{RT} model and the BS^{OS} model. A negative value indicates that inflation has been underestimated.

Outline

1 Introduction

2 Literature review

3 Methodology

4 Results

5 Conclusions

Main takeaways

1. Pure estimates of online price changes is already effective in nowcasting food inflation.
2. Incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
3. This approach outperforms a variety of frameworks, including judgemental methods.
4. Marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
5. During COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
6. Meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

NBP Narodowy Bank Polski

Main takeaways

1. Pure estimates of online price changes is already effective in nowcasting food inflation.
2. Incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
3. This approach outperforms a variety of frameworks, including judgemental methods.
4. Marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
5. During COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
6. Meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

NBP Narodowy Bank Polski

Main takeaways

1. Pure estimates of online price changes is already effective in nowcasting food inflation.
2. Incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
3. This approach outperforms a variety of frameworks, including judgemental methods.
4. Marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
5. During COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
6. Meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

NBP Narodowy Bank Polski

Main takeaways

1. Pure estimates of online price changes is already effective in nowcasting food inflation.
2. Incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
3. This approach outperforms a variety of frameworks, including judgemental methods.
4. Marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
5. During COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
6. Meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

NBP Narodowy Bank Polski

Main takeaways

1. Pure estimates of online price changes is already effective in nowcasting food inflation.
2. Incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
3. This approach outperforms a variety of frameworks, including judgemental methods.
4. Marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
5. During COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
6. Meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

Main takeaways

1. Pure estimates of online price changes is already effective in nowcasting food inflation.
2. Incorporating information on online prices into model-based frameworks delivers a substantial increase in the nowcast accuracy.
3. This approach outperforms a variety of frameworks, including judgemental methods.
4. Marked improvement can be obtained for a number of inflation components, especially those experiencing high volatility throughout the year.
5. During COVID-19 the nowcasting quality relatively improved and remained comparable with judgemental nowcasts.
6. Meticulous product selection and expenditure weighing is essential for providing accurate both in-sample fit as well as out-of-sample nowcasts.

NBP Narodowy Bank Polski