Business tendency survey analysis using artificial intelligence methods

Michał Bernardelli

IV Kongres Statystyki Polskiej
Warszawa, 2nd-4th July, 2024

Presentation outline

SGH

www.sgh.waw.pl

Goal and motivation

Use AI algorithms to increase prediction accuracy and discover hidden data patterns

Assess suitability of the long short-term memory model for forecasting economic time series

Check usefulness of the business tendency survey in predicting price indices of the industry's sold production in Poland

AI approach to forecasting time series

AI areas of applications

- Data exploration
- Process automation
- Optimisation
- Decision-making assistance

Examples

- Credit assessment
- Algorithmic trading
- Accounting and auditing automation
- Fraud detection
- Risk management
- Portfolio optimisation

Time series

- N-BEATS model
- Prophet algorithm
- Time Series Transformer (TST)
- PatchTST and PatchTSMixer models
- Long short-term memory (LSTM) model

SGH

Long short-term memory (LSTM) model
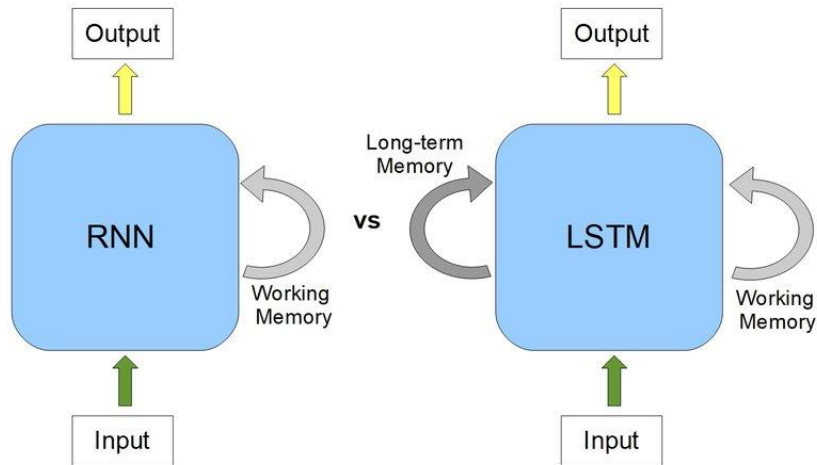
History

Introduced by Sepp Hochreiter and Jürgen Schmidhuber in 1997

Idea

Recurrent neural network (RNN) remembering information from the distant past while considering new input data

Application

Exploit by Google, Amazon, Apple, Microsoft, Facebook

Source: ResearchGate

SGH

LSTM model - applictions

Robot control	Recognising patterns	Time series prediction	Speech recognition	Rhythm learning	Hydrological rainfall–runoff modeling
Music composition	Grammar learning	Handwriting recognition	Human action recognition	Sign language translation	Protein homology detection
Predicting subcellular localization of proteins	Time series anomaly detection	Several prediction tasks in the area of business process management	Prediction in medical care pathways	Semantic parsing	Object co-segmentation
Airport passenger management	Short-term traffic forecast	Drug design	Market Prediction	Activity Classification in Video	Image recognition

SGH

Data characteristics (explained variable)
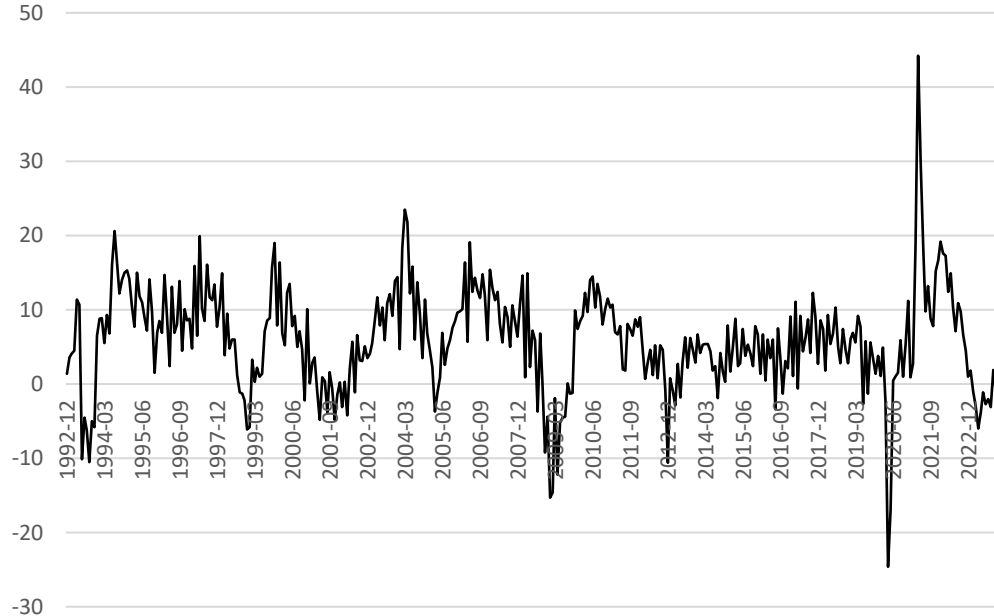
Monthly price indices of sold production of industry in Poland from December 1992 to February 2024

Significant change of pattern since 2019.

Source: Own computation based on Central Statistical Office data

SGH

Data characteristics (explanatory variables)

Monthly business tendency survey in the Polish manufacturing industry conducted by the Research Institute for Economic Development at SGH Warsaw School of Economics

Respondents in this survey evaluate changes in various economic areas by answering eight questions about:
- production volume
- total and export orders
- finished goods inventory
- selling prices
- employment
- financial standing
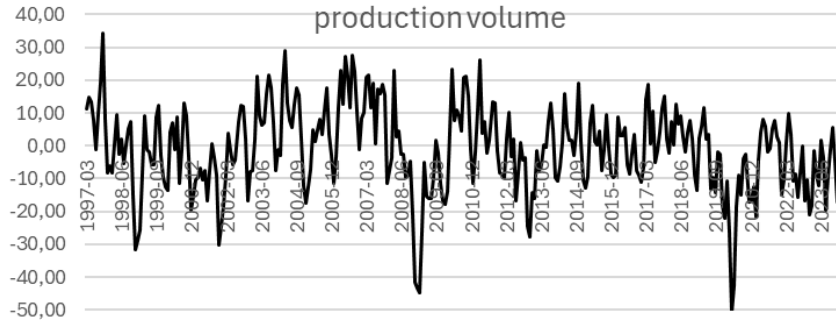- general economic situation in Poland

Two versions: retrospective about the current and perspective about the future

Based on the responses, the balance is calculated as the difference between the percentages of positive and negative answers

SGH

Data characteristics (explanatory variables)

production volume

employment

financial standing

Balances of responses to the selected questions from the RIED business tendency survey.

Answers to different questions are not the same (questions carry significantly different knowledge)
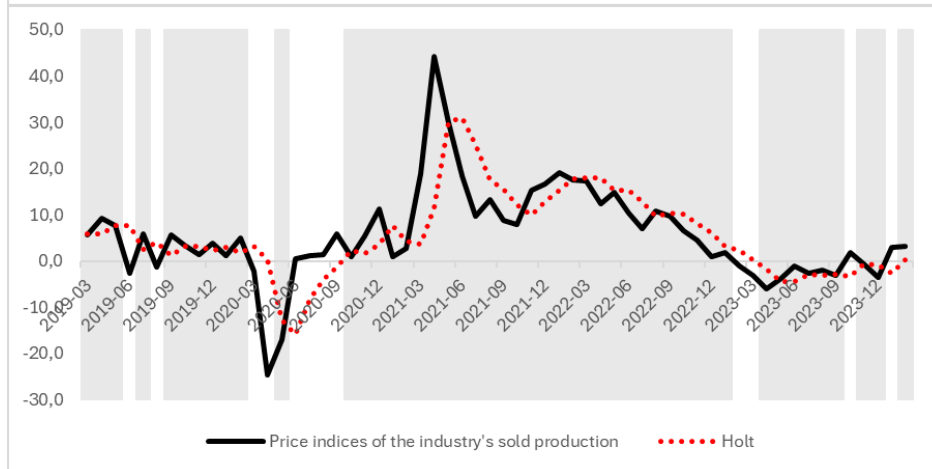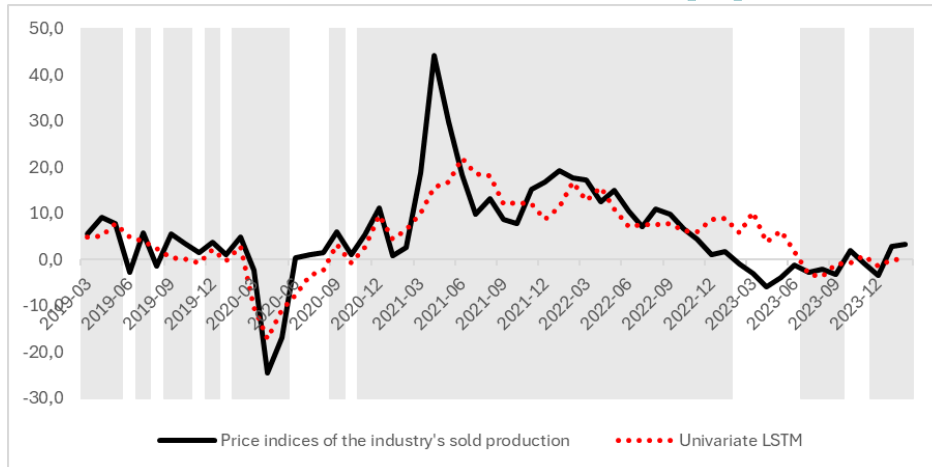
Empirical analysis

Forecasting using a single time series, which is both an explained variable and an explanatory variable.

Prediction of the price indices of sold production of industry time series in Poland based on multivariate LSTM model and the set of all answers from the business tendency survey.

LSTM – univariate approach

Test set: each month starting from March 2019

Forecast horizon: one month

Parameters: dropout rate 0%, 50 epochs

Window length for the training set: 12 months

- too long, gives too little weight to recent observations, and does not consider changes in the economic situation over the years

- too short, does not provide historical knowledge about the behaviour of the analysed time series

Comparison to the reference forecast done with double exponential smoothing (Holt linear)

The months in which the predictions' sign coincided with the actual sign of the price index are marked with a grey background

www.sgh.waw.pl

LSTM – multivariate approach

Price indices of the industry's sold production .... Multivariate LSTM
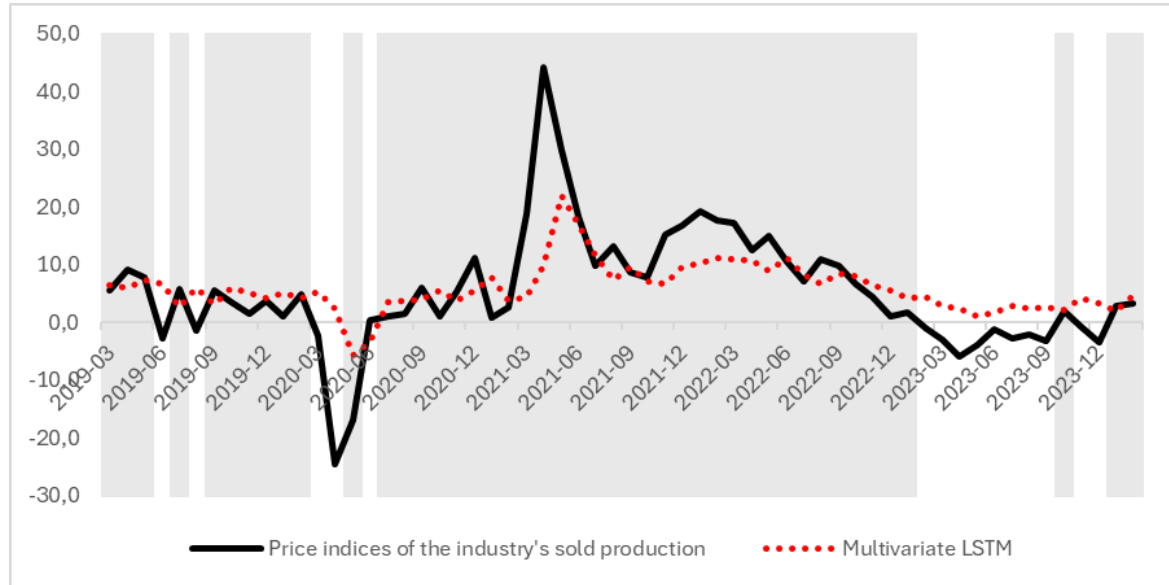
Test set: each month starting from March 2019

Forecast horizon: one month

Window length for the training set: no limitation

The months in which the predictions' sign coincided with the actual sign of the price index are marked with a grey background

- Possibility to use on input many time series simultaneously

- Not using the explained time series as the explanatory variable

- Connection between respondents' opinions and hard data from the economy: identification of periods and conditions when respondents indicate the correct direction of changes and moments when they differ from the actual macroeconomic data.

SGH

Results

Period	Statistics		MAPE		
	Mean	SD	Holt	Univariate LSTM	Multivariate LSTM
Whole test period	5.48	10.06	1.75	1.53	1.49
2019.03 - 2020.02	3.72	3.53	1.16	1.00	1.50
2020.03 - 2021.02	-1.13	9.92	5.04	2.94	2.52
2021.03 - 2022.02	17.10	10.57	0.53	0.50	0.37
2022.03 - 2023.02	8.72	6.05	0.96	1.58	1.07
2023.03 - 2024.02	-1.45	2.75	1.18	1.99	2.21

- The MAPE error was the smallest for the multivariate approach during the entire period under consideration. It is also confirmed in two of the five annual test set periods.
- The multivariate LSTM model performs worse in periods with small change amplitudes than the reference Holt method or the univariate LSTM model.
- The direction of changes recognised by the LSTM model coincided with 75% of the actual changes.
- Errors appeared mainly in recent years, which are periods of relative stability (entrepreneurs are slightly more optimistic about the future than the statistical data suggests).

SGH

Concluding remarks

- Holt method often gives predicted values as just a shift in historical values

- LSTM model performs better during less predictable periods

- LSTM model is suitable for analysing single time series with not very large (monthly) time granularity

- Close relationship between the opinions expressed by respondents in the monthly business tendency survey in the Polish manufacturing industry and hard data from the economy

- On average, a multivariate approach was slightly more effective (mismatch is not the model's fault but rather the respondents' incorrect assessments)

SGH

Contact

dr hab. Michał Bernardelli, prof. SGH

michal.bernardelli@sgh.waw.pl

Thank you for your attention!