# Forecasting regional banknote withdrawals and lodgements

Benedikt Sonnleitner | Fraunhofer IIS and Vives University of applied Sciences
Jelena Stapf | Deutsche Bundesbank
Kai Wulff | Deutsche Bundesbank

## Abstract

For Central Banks it is important to ensure the availability of cash to the economy. Forecasting the demand for cash on a granular level is crucial in the process to keep logistic costs low, while being resilient to demand or supply shocks. We benchmark statistical and machine learning methods on forecasting demand and supply of cash, using data on transactions of six regional branches of Deutsche Bundesbank. We show that elaborate forecasting methods can substantially improve forecasting accuracy and inventory performance for this use-case.

---

---

# Introduction

Cash is the most frequently used means of payment at the point of sale in Germany and its stable supply is central for the trust of citizens in the monetary system. To ensure this uninterrupted supply, the Bundesbank operates a nationwide network of 31 branches. Accordingly, in all of the Bundesbank branches, a substantial stock of cash is held. Furthermore, to enable smooth and stable supply of banknotes for each denomination, transport and inventory planning relies on forecasted regional cash demand. In our study, we investigate whether machine learning methods from business forecasting improve forecast accuracy and reduce inventory cost compared to a simple benchmark. Figure 1 depicts the supply chain of banknotes starting with the Bundesbank Logistics Centre which supplies regional branches. These deliver banknotes to and receive lodgements from their local customers.

**Figure 1. Schematic supply chain of bank notes**
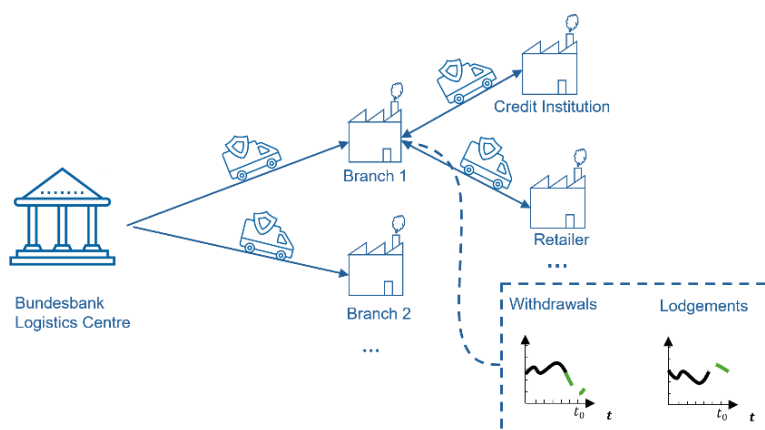


The current planning process uses a weekly forecast of withdrawals and lodgements of regional branches adjusted by seasonal factors which is done implicitly manually. Based on these, the Bundesbank Logistics Centre plans the transports for filling or disposal. Thereby it considers various constraints such as deliveries from printing works, cross border and international transports and the schedule of the accompanying regional police forces.

There are two related use-cases for cash demand forecasting: First, the issuers of legal tender currency make yearly or longer-term projections for production requirements. These encompass break downs for the denominational split of the banknotes to be produced and incorporates replacement ratios for so called unfit banknotes, i.e. torn and soiled banknotes which cannot be paid out any longer. For example Miller (2017) and Bartzsch et al. (2023) use a broad range of structural time series models. Cash management on an operative level requires shorter forecast horizons (21 days vs. several months), and spatially more granular forecasts (local branches vs. whole countries).

Second, a strand of literature evolves around replenishment optimization problems for ATMs. This has been the foundation of the NN5 competition (Crone, 2008), whose data set is since used in several forecasting studies. Forecasting for local central bank branches differs, since they deliver cash to commercial banks instead of consumers. This introduces other demand patterns and adds an additional layer of forecast complexity compared to ATM-forecasting by incorporating features from long-term demand forecasting such as recycling and replacement ratio issues.

# The study

In order to facilitate a more formalized approach to forecasting regional cash demand for banknotes, our study employs anonymized daily transaction data for six regional branches of the Deutsche Bundesbank. Beginning in January 2017, the interval encompasses more than five years of daily data per denomination and branch.

Bundesbank branches might generate out-payable banknotes by processing paid-in banknotes on the other day. The difference between processed banknotes and banknotes that must be destroyed because of soil and stain issues, i. e. the replacement ratio, influences the stock of out-payable banknotes and therefore delays replenishment. As the net demand of banknotes cannot be directly observed, we combine historical data of cash orders with data of lodgements and banknote processing, resulting in four individual datasets.

The sheer amount of time series prevents manual forecast modeling for each single time series, and there are more data points available than typically on high-level data. Accordingly, the forecasting task differs substantially from macro-economic modeling that is typically done in central banks and is closer to business forecasting where recently machine learning has blossomed.

Statistical forecast models assume that the time series follow some pre-defined structure. Most prevalent, exponential smoothing and ARIMA models have been shown to work well for many business forecasting tasks (Hyndman, 2021). By contrast, machine learning methods impose no, or few assumptions about the time series structure. This allows them to learn patterns purely from the data, however makes them more data-hungry in training. Accordingly their potential can only be used if sufficient data is available. Commonly, these networks are trained across all time series in a data set, instead of training one model per time series as it is typically done for the mentioned statistical forecast methods. The most prevalent machine learning approaches are on one hand decision tree based methods, such as Random Forests, XgBoost, and LightGBM (Makridakis, Spiliotis, & Assimakopoulos, 2020; Januschowski, et al., 2022). On the other hand various neural networks, with a dynamic development of new architectures evolved (Smyl, 2020; Salinas, Flunkert, Gasthaus, & Januschowski, 2020).

Historically there was a broad consensus that simple (statistical) forecasting methods should be preferred over complex ones (Makridakis & Hibon, 2000). However in recent years evidence has intensified that machine learning models can improve accuracy over statistical approaches, as these won several forecasting competitions and are used increasingly in industry (Makridakis, Spiliotis, & Assimakopoulos, 2020; Bojer & Meldgaard, 2021; Schaer, Svetunkov, Yusupova, & Fildes, 2022). Nevertheless, whether machine learning improves performance over statistical forecasting, depends on the specific architectures and data sets (Petropoulos, Makridakis, Assimakopoulos, & Nikolopoulos, 2014; Godahewa, Bergmeir, Webb, Hyndman, & Montero-Manso, 2021; Hewamalage, Bergmeir, & Bandara, 2021).

We compare well established statistical demand forecasting approaches (ETS and ARIMA(X)) and neural networks (MLP and DeepAR) specifically designed for demand forecasting against the seasonal naïve benchmark, a simple method that mimics the current manual forecast. We evaluate each forecasting method in terms of forecast accuracy, and in terms of quality of decisions. For the latter, we simulate the inventory decisions if we had planned upon the respective forecast method and calculate the induced costs. We do these simulations across different target service levels.

## Results

We find that all evaluated business forecasting methods in most cases improve forecast performance over the seasonal naïve benchmark. Figure 2 summarizes the forecast accuracy in different aspects for the used data set. Each curve and symbol marks one forecast method. The labels on the x-axis are different performance measures, and their value is shown on the y-axis. We report the relative performance in comparison to the naïve benchmark (i.e 1 is equally accurate). The lower a symbol, the better performs the respective forecast method. We find that:

- All elaborate methods improve forecast bias, that means they have less systematic over- and underforecasting, than the seasonal naïve forecast. The two evaluated machine learning methods exhibit the smallest forecast bias, measured as relative Absolute Mean Error (rAME).

- Apart from ARIMAX, all elaborate methods improve on mean point forecast accuracy. DeepAR, a neural network specifically designed for demand forecasting performs best. This is measures by the relative Mean Absolute Error (rMAE), and the relative Mean Squared error (rMSE).

- Apart from ARIMAX, all elaborate methods improve on the quantile forecast accuracy, where additional to the mean point forecast, an uncertainty estimate is required. DeepAR outperforms the other methods on the 90%-quantile but is worse on the 99%-quantile. This is measured by the relative PINball loss (rPIN).

For an in-depth description of the used performance measures and evaluation scheme, please see the related paper (Sonnleitner, Stapf & Wulff, 2024).

**Figure 2. Forecast accuracy benchmark**



The improvements in forecast accuracy translate into reduced inventory costs. Figure 3 shows the inventory performance of the different evaluated methods. Each solid-colored line corresponds to one forecasting method and shows the relation between held average stock and the achieved service level. The farther left top a method is, the better is the respective inventory performance, as with the same held inventory a higher service level is reached. The black, dotted vertical lines show the deviation between achieved service level and target service level. If they are below the respective line, the target service level is exceeded, if they are above, the method fails to reach the target service level.

All elaborate methods improve the inventory management substantially over the seasonal naïve benchmark, and DeepAR still ranks best.

**Figure 3. Inventory performance benchmark**
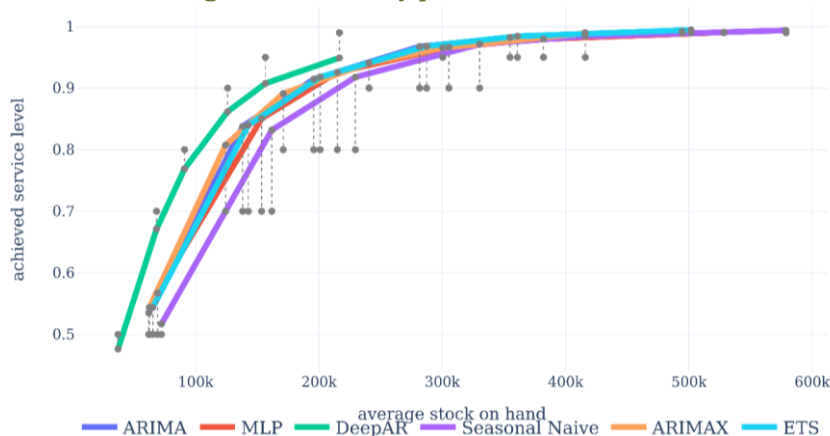
## Suggestions and Limitations

Our study suggests that central bank logistics for regional cash supply might benefit from machine learning based forecasts. These can improve forecasting performance and thus provide a better support for planning. More granular forecasts, in our example per branch and denomination on a daily level are supported by the existence of many related time series with comparably many observations. In this setting machine learning methods score high forecasting results.

Our analysis focuses on forecasting and inventory management in "standard times". However well-functioning operations in central banks are especially crucial in times of crisis. In this respect a so called 100% service level target is formulated by central banks. Yet the fulfillment of these target level is rather ensured by large central reserves and strategic stocks than solely by regional transport operations. Further, in our study we assume a simple policy for inventory decisions, that does not fully mimic the additional constraints encountered by cash-in-transit operations. Nevertheless, we hope that our results provide a starting point to improve forecasting for logistic purposes in central banks.

## References

Bartzsch, N., M. Brandi, R. de Pastor, L. Devigne, G. Maddaloni, D. P. Restrepo, and G. Sene (2023). Forecasting banknote circulation during the covid-19 pandemic using structural time series models. *Deutsche Bundesbank, Discussion Paper (20)*.

Bojer, C. S., & Meldgaard, J. P. (2021). Kaggle forecasting competitions: An overlooked learning opportunity. *International Journal of Forecasting*.

Crone, S. F. (2008). *NN5 competition*. Retrieved from http://www.neural-forecasting-competition.com/NN5/

Godahewa, R., Bergmeir, C., Webb, G. I., Hyndman, R. J., & Montero-Manso, P. (2021). Monash Time Series Forecasting Archive. *Neural Information Processing Systems Track 2021.*

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *International Journal of Forecasting*.

Hyndman, R. J. (2021). *Forecasting: principles and practice.* OTexts: Melbourne, Australia.

Januschowski, T., Wang, Y., Torkkola, K., Erkkilä, T., Hasson, H., & Gasthaus, J. (2022). Forecasting with trees. *International Journal of Forecasting*.

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M5 Accuracy competition: Results, findings and conclusions. *International Journal of Forecasting*.

Miller, C. (2017). Addressing the limitations of forecasting banknote demand. *International Cash Conference 2017.* Deutsche Bundesbank.

Petropoulos, F., Makridakis, S., Assimakopoulos, V., & Nikolopoulos, K. (2014). 'Horses for Courses' in demand forecasting. *European Journal of Operational Research (European Journal of Operational Research)*.

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*.

Schaer, O., Svetunkov, I., Yusupova, A., & Fildes, R. (2022). Survey: Forecasting software trends in a challenging world. *Institute for Operations Research*.

Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*.

Sonnleitner, B., J. Stapf & K. Wulff (2024), Benchmarking short term forecasts of regional banknote lodgements and withdrawals. Deutsche Bundesbank, Discussion Paper (39).

## About the author(s)

**Benedikt Sonnleitner** is researcher at the Vives University of applied Sciences in Belgium and Senior Scientist at the Fraunhofer IIS in Germany. His research focuses on demand forecasting in operations and especially the interface between forecasting and decision making.

**Jelena Stapf** heads a division in the Cash Department of Deutsche Bundesbank and is responsible for branch operations, IT processes and cash automation. Research topics comprise RNDs from inflation-linked options, price discovery on traded inflation expectations, stock market volatility and term structure issues.

**Kai Wulff** is Senior Expert in the Cash Department of Deutsche Bundesbank, responsible for various cash related IT projects as Project Manager. His work focuses on a wide range of analytical topics, data-driven issues and data governance aspects.