Time series editing using ML

Emma Stavås, Statistics Sweden, emma.stavas@scb.se

Martin Odencrants, Statistics Sweden, martin.odencrants@scb.se

Jens Malmros, Statistics Sweden, jens.malmros@scb.se

Abstract

A Variational Autoencoder (VAE) is applied to aid validation of financial time series in the survey Monetary financial institutions, balance sheet statistics. Every production cycle involves validation of tens of thousands of series making traditional rule-based editing and validation procedures in the current production system infeasible. The VAE learns a latent representation based on validated time series from previous production cycles. A time series is identified as erroneous when the reconstruction error exceeds a predefined threshold in combination with the level of the reported series. Preliminary results show that the VAE identifies unusual reporting patterns, offering a complementary signal to existing rule-based checks.

The use of ML in statistical production was first highlighted in the UNECE HLG-MOS Machine Learning Project (2019), where editing was identified as a key use case. More recently, the AI/ML for Official Statistics initiative has introduced a dedicated work package on ML-based validation and editing, focusing on how AI/ML can enhance the data editing and imputation process within statistical offices. The use of a VAE in the current use case demonstrates how unsupervised learning can complement existing validation procedures in the current production system.

Keywords: Machine learning, editing, variational autoencoder

1. Introduction

a. Background

In recent years, Statistics Sweden has undertaken a strategic review of its data editing and validation practices, with the objective of reducing manual micro-editing in surveys based on direct data collection. Initiated in 2021, this review resulted in a redesigned editing process and an implementation plan aimed at increasing efficiency while maintaining quality in statistical output. Central to the new process is the principle that validation efforts should be selective and focused on identifying errors that have a significant impact on the overall quality of the published statistics. The review further emphasised the use of automated procedures rather than manual reviews.

Against this background, efficient editing and validation procedures are central components of the statistical production process, particularly in high-frequency surveys where data quality must be ensured within short production time frames. In financial statistics, such as the Monetary Financial Institutions (MFI) assets and liabilities survey, each production cycle involves the validation of tens of thousands of time series. This volume makes it infeasible to rely solely on traditional rule-based editing systems, which typically are limited in their ability to account for previously observed reporting patterns. Consequently, there is a need for procedures that better filter and prioritise observations according to their relevance for overall data quality, while maintaining a high level of precision in the identification of implausible values.

In this context, this paper presents a use case where an unsupervised machine learning approach is introduced as a complementary editing tool. Inspired by similar work at the European Central Bank (ECB) (Fajt Mayer, Raimondo Quaratino, & Javier Cordero, 2024), a Variational Autoencoder is used to support the editing of financial time series in the MFI assets and liabilities survey.

2. VAE for anomaly detection

Labelled examples of reporting errors require costly manual annotation, limiting the applicability of supervised machine learning methods. Previous work has highlighted the potential of unsupervised approaches for supporting editing tasks and the prioritization of observations for review (UNECE, 2021, ss. 39-48). Building on this idea, we formulate the MFI use case as an anomaly detection task.

The method adopted is a Variational Autoencoder (VAE) (Kingma & Welling, 2013). A VAE is a probabilistic neural network consisting of an encoder and a decoder. The encoder maps input data to a lower-dimensional latent representation, while the decoder reconstructs the original input from this representation. Reconstruction errors have been widely used for VAE-based anomaly detection, including applications to multivariate time series data (An & Cho, 2015) (Xu, o.a., 2018). Observations associated with unusually large reconstruction errors are treated as anomalies.

For the MFI use case, the VAE is trained on time series validated in previous production cycles to learn typical reporting patterns. New reported values are evaluated using the squared reconstruction error between reported and reconstructed values. Reported values with large reconstruction errors are treated as potential anomalies that may correspond to implausible or misreported values. The corresponding time series are flagged and prioritized for review by subject-matter statisticians. Percentile-based thresholds are applied to ensure a manageable number of flagged time series.

3. Data, model specification and implementation

a. MFI data

The statistics for this use case cover assets and liabilities of MFIs, which include banks, housing credit institutions and finance companies. The statistics are based on a census survey where most MFIs report monthly, while some report

quarterly or annually. The primary data source for this work is the time series data from the survey responses.

The time series are defined by reporting institution and variable code. The variable codes are structured into dimensions separated by underscores, where each dimension carries information about the reported value. The reported series vary substantially in magnitude, mainly due to differences across reporting institutions and variable codes.

b. Data cleaning, preprocessing and feature engineering

i. From the original 180 months we kept the most recent 120 to ensure relevance while maintaining sufficient training data.

ii. Stock variables were differenced to obtain flows, ensuring consistency across variables.

iii. Zero imputation was chosen to address missing values for simplicity and transparency. True zeros were considered rare.

iv. Series containing only zeros were excluded and variable codes no longer in use were removed. Series with only zeros in the last 24 months were excluded.

v. A chronological train-test split was used: 108 months for training and the last 12 months for testing. The final training set included 189 reporters and 3 516 unique variable codes.

vi. All-zero series were removed again from training data to handle variable codes or reporters appearing only in the test set.

The time series were normalized using a MinMax scaler with an additional 5% buffer around the training range. The normalized series were transformed into overlapping rolling windows of 12 months. Variable codes were decomposed into 11 components (e.g. account, maturity, currency, counterparty) and combined with the reporter identifier. These categorical variables were one-hot

encoded and then compressed using an autoencoder to a low-dimensional representation, resulting in two additional features.

c. Model specification

The encoder and decoder were composed of fully connected neural network layers. The number of layers, latent dimensions, hidden units, dropout rate, learning rate and batch size were treated as hyperparameters and optimized during hyperparameter search.

The model was trained to reconstruct the 12-month input window, and the final observation, corresponding to the last reported value, was retained for anomaly detection. With y_{it} denoting the reported value of series i in period t and $\hat{y}_{it}$ the corresponding reconstructed value, the squared reconstruction error was calculated as

$$e_{it} = (y_{it} - \hat{y}_{it})^2$$

The reconstruction error was squared to emphasize larger deviations between reported and reconstructed values. Separate thresholds were constructed for growth-rate and non-growth-rate series. The thresholds were defined at a set percentile of the empirical distribution of squared reconstruction errors from a holdout period. During inference, a reported value is flagged when its squared reconstruction error exceeds the corresponding threshold value.

4. Results

Candidate models were evaluated by comparing the reconstruction errors and the model with the lowest validation reconstruction error was chosen. The complete anomaly detection system was evaluated by partially manual annotation across a set of reporters with different characteristics. The anomaly detection threshold was set at the 99.5th percentile to obtain an operationally

manageable number of alerts. More extensive testing is required to fully assess the model's strengths and limitations.

An initial version of the anomaly detection system has been implemented in the production system for further evaluation by subject-matter statisticians. The results are promising, although not yet fully satisfactory. For example, we observed that the model tends to reconstruct highly volatile series less well. To mitigate this issue, two additional filters were introduced to prevent low-volume series and series exhibiting only minor changes relative to the previous period from being flagged.

5. Discussion

The preliminary results indicate that the method can be used to prioritise reported values for review. The approach is consistent with the revised editing process at Statistics Sweden, where efforts are directed towards observations that are most relevant for the quality of the published statistics. The VAE anomaly detection system supports a human-in-the-loop editing process by identifying reported values that may require further review by subject-matter statisticians. At the same time, more extensive evaluation is needed to assess its impact on editing efficiency.

We would like to investigate the reconstruction error normalisation proposed by the ECB. By taking the volume of the underlying series into account, it may be possible to further improve the prioritisation of reported values for review. Another area for future work is the establishment of feedback mechanisms from production and monitoring of model performance.

This project has contributed to the ongoing development of machine learning capabilities at Statistics Sweden. As part of a broader effort to establish MLOps-based ways of working, the project provides practical experience of developing, deploying and maintaining machine learning models in a statistical production environment.

6. List of references

An, J., & Cho, S. (2015). Variational Autoencoder based Anomaly Detection using Reconstruction Probability.

Fajt Mayer, T., Raimondo Quaratino, G., & Javier Cordero, J. F. (2024). Variational autoencoders for multivariate time-series outlier detection. *12th biennial IFC Conference: "Statistics and beyond: new data for decision making in central banks" .*

Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. *Proceedings of the International Conference on Learning Representations.* doi:https://doi.org/10.48550/arXiv.1312.6114

United Nations Economic Commission for Europe. (2021). *Machine Learning for Official Statistics.* United Nations, Geneva. Retrieved from https://unece.org/sites/default/files/2022-09/ECECESSTAT20216.pdf

Xu, H., Feng, Y., Chen, J., Wang, Z., Qiao, H., Chen, W., . . . Pei, D. (2018). Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications. *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18* (pp. 187-196). Lyon, France: ACM Press. doi:10.1145/3178876.3185996