Department of Economics

University of Reading

A Pound Centric look at the Pound vs. Krona Exchange Rate Movement from 1844 to 1965

by Andrew Clark

A Pound Centric look at the Pound vs.Krona Exchange Rate Movement from 1844 to 1965

Andrew Clark[*][†]

October 2020

Abstract

A longitudinal (1844-1965) study of the Pound Krona exchange rate is conducted utilizing London Times article news sentiment, gold price, GDP, and other relevant metrics to create a dynamic systems state-based model to predict the Pound Krona yearly exchange rate. The created model slightly outperforms a naive random walk forecasting model.

JEL classification codes: C32, C53, C63, E17, F31

Keywords: Econometrics Machine Learning Dynamic Systems Complex Systems

1 Introduction

Traditional econometrics models have a hard time beating random walk models [1]. The forecasting accuracy of economics models is often poor, and focuses on unilateral signals or VAR type autoregressive models. Two recent trends in economics are utilized in this paper to explore alternatives for increasing the accuracy of forecasting models, machine learning and dynamic systems modeling.

With the rise of big data and machine learning, interest in the application of machine learning to economics is increasing, with many central banks and research organizations exploring the applicability of machine learning to some of their economic workflows. Machine learning can be utilized for regression type prediction problems, sentiment analysis, and many other applications. Machine learning is just an extension of statistical modeling, with its goals often different than statistical modeling. Separating the signal from noise can sometimes be difficult, with the general consensus being that machine learning can augment existing methods and cause incremental improvement, vs the revolution some had anticipated.

Complexity science and dynamical systems modeling have risen in prominence in the field of Economics in the aftermath of the financial crisis, with some university creating research groups on complexity and network science. The modeling paradigms for complexity science and dynamical systems have not been standardized or are as accessible as in econometrics and machine learning modeling.

The purpose of this paper is to:

1. Illustrate the application of machine learning for sentiment analysis of historical news paper articles.

2. Create a machine learning model for forecasting yearly exchange rates that is generalizable, but also optimized for the time period and the Krona/-Pound.

3. Illustrate the application of dynamical systems modeling as a paradigm for embedding machine learning or econometric models.

2 Literature Review

In the last 5 years or so, Machine Learning and Sentiment Analysis have begun to be an economic research area, with a variety of papers appearing from many different sources. A noteworthy paper from the Bank of England has described Machine Learning and how it is applicable to the economics profession in a very easy to interpret way[2].In their introductory paper, the authors present a critique of machine learning that is primarily focuses on prediction, while many policy applications of economics - and econometrics by extension - are primarily focused on causal inference. Reference around how to detail handle this potential problem, along with the "black box" nature of many machine learning algorithms, the researchers note how many policy problems can be divided into prediction and causal inference parts, allowing the use of machine learning for the predictive aspect, while it is still an open research question on how add causality to machine learning[3][4]. This paper is focused on the application of correlation modeling not casual.

2.1 Economic Sentiment Analysis

Diving into Sentiment Analysis for Economics, there is an excellent paper by that provides an overview of the uses of text mining for central banks[5]. Citing heavily from the established natural language processing literature, the handbook provides and excellent introduction to the possibilities of text mining economics, which is applicable to this paper. Primarily focused on unsupervised machine learning techniques, the handbook also provides a brief introduction to the uses of supervised learning in text analysis. As an integral part of central banking is trust, text mining allows the central bank to monitor the sentiment of the financial markets in their jurisdiction, and/or affect the jurisdiction of the

central bank to provide insight into investors mentality. Text analysis also allows the central bank to evaluate if their messaging is being accurately understand by the public.

A paper by Shapiro, Adam Hale, Moritz Sudhof, Daniel Wilson uses text analysis of economic and financial newspaper articles to measure economic sentiment [6]. 16 US newspapers were used to create "corpus" of economic news articles to be aggregated into a monthly sentiment index. The monthly news sentiment index was then used to ascertain if/how news sentiment affects economic variables. For modeling purposes, the federal funds rate, PCE, S&P 500 stock price index, etc. were used as "hard" data inputs into a local projections model. A slight lift was found from the inclusion of the news sentiment measures in a model the forecasting performance of economic variables, such as PCE. The proven hypothesis that news affects economic variables is a key part of the modeling done in this paper.

2.2 Complex Systems Theory

In order to model the Pound/Krona exchange rate movement over the time frame studied, a novel approach of a dynamical complex systems model will be created, utilizing a recently open-sourced computer automated design tool called cadCAD[7]. cadCAD (complex adaptive systems computer-aided design) is a python based, unified modeling framework for dynamical systems and differential equation simulations created by BlockScience. It is capable of modeling systems at all levels of abstraction from Agent Based Modeling (ABM) to System Dynamics (SD) with integration of with existing data science workflows and paradigms.[7]

Complex systems are systems that combine many different elements that interact with each other in unique and complex ways. Complex system behavior isn't easily modeled due to properties that include nonlinearity, adaption, feedback loops, emergence and spontaneous order[8]. Complex systems theory is very interdisciplinary, with heavy influenced from economics, engineering, physics, biology, to name a few, in an attempt to holistically understand the system. Approaching economic forecasting in a complex systems sense allows for a novel approach with the hypothesis that taking the whole is greater than the sum of the parts will produce higher predictive accuracy than traditional reductionist approaches to econometric modeling.

Economics is undergoing a transformation from a "Physic's envy" reductionist approach to begin to embrace complex systems theory and its potential to add a more holistic approach to economic modeling[9]. As the paper by Focardi outlines, Artificial Economics and state based modeling is gaining popularity and is an open research area, which this paper contributes to.

3

3   Problem Framing

The world changed drastically between the mid 19th century and the mid 20th century, the world's central banks moved away from the gold standard, two world wars changed the face of Europe, and the world's economies began using floating exchange rates. Throughout this immense amount of change and upheaval, two central banks and their governments took impeccable records of the changing events, the Bank of England and the Sveriges Riksbank. Reporting the news during this volatile period, The Times has preserved archives that can be tapped upon to take a look into the psyche of the English people. Hardly a better time period could be chosen to see how major events affected both news paper sentiment and exchange rates. England was embroiled in both world wars while Sweden stayed neutral during both. The above-mentioned facts make the forecasting of exchange rates between these nation-states during the specified period a wonderful backdrop for creating a model and testing a complex model.

In the following section, we will go through the data acquisition and preprocessing before discussing the model construction.

4   Data Aggregation, Preparation and Cleansing

4.1   News Articles

In order to obtain historical news articles mentioning the Pound or Krona, the Gale Times Digital Archive was used[10]. The Gale Times digital archive contains optical character recognition (OCR) software to turn images into text to allow news articles to be downloaded as text files. Advanced search criteria is available in the Gale archive, and the following search criteria were used:

1. Title = krona or pound or sterling

2. Publication Date > 1843 < 1966

3. Publication Section= "Business News" OR "News"

4. Document Type= "Article" OR "Editorial"

As no bulk download option or application program interface (API) was available on the database, 670 articles that met the before mentioned criteria were manually downloaded. The news articles were then imported into a Python 3 Jupyter notebook and cleaned by removing a Gale imposed disclaimer on OCR. Additional text pre-processing was conducted, such as removing line returns, converting all characters to lower case, and removing stop words. This preprocessed data was downloaded as a csv for use in our sentiment analysis model described below.

4.2   Economic Data

Exchange rate data was obtained from the Sveriges Riksbank's detailed historical archives of the bank[11][12]. The UK Gold data was obtained from MeasuringWorth[13], the Swedish GDP obtained from the Riksbank[14] and the UK GDP was obtained from FRED [15]. All of the data was preprocessed by renaming columns, changing datatype formats, etc, subset to include only the relevant time range, 1844-1965, and merged together into a format usable in the simulations.

4.3   News Articles

Although we will use an actual natural language processing algorithm for the sentiment modeling, we are using a word list for analysis of the news articles. The Loughran-McDonald Sentiment Word Lists were used to construct a word count sentiment index of all of the articles downloaded[16]. The index was created by aggregating all news articles for a given year, counting the individual words for number of occurrence, and comparing the overlap with the Loughran-McDonald Sentiment Word Positive, Negative and Uncertainty word lists. The index was calculated by positive, negative, and uncertainty divided by total words, respectively. The following algorithm was used to determine the sign given to the data:

$$\text{if negativePercentage} > \text{positivePercentage and negativePercentage} > 0.51 :$$

$$\text{sign} = \text{NEGATIVE}$$

$$\text{elif positivePercentage} > \text{negativePercentage and positivePercentage} > 0.51 :$$

$$\text{sign} = \text{POSITIVE}$$

$$\text{else:}$$

$$\text{sign} = \text{NEUTRAL}$$

For purposes of our modeling, we will use a more sophisticated sentiment modeling solution, although insight is provided by examining a traditional word counting approach. From figure 1, we can see that the majority of articles had a neutral sentiment, with a significant amount of negativity from 1920 to 1960, with a low point of 1940-1945 with extremely low values. Due to the historical context and England's precarious situation, extreme negative values makes sense.

5   Modeling

5.1   Sentiment Analysis

Sentiment Analysis is a very nuanced field of study that requires vast amounts of training data, a deep knowledge of the target domain, as well as insight into the linguistic meanings of words during the period of study.
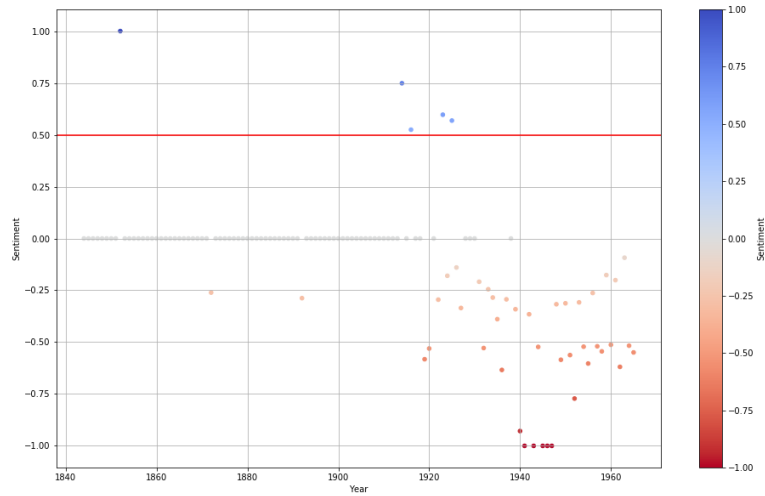
Figure 1: Loughran-McDonald Sentiment Word Lists analysis. -1 means very negative sentiment and 1 means very positive sentiment.

To examine historical documents that first must go through optical character recognition (OCR) to turn images into text, only exacerbates the problem, see section 4. As to the Author's knowledge, no extensive sentiment classification datasets have been created for English Financial Newspapers between 1844 and 1965. As such, a few short cut optimizations will be made in order to complete this study. After extensive research, the Flair library by Zalando Research was chosen due to its pretrained word embeddings, model structure, and that it was built in PyTorch, a leading Python based deep learning library[17]. Flair uses cutting-edge natural language processing (NLP) techniques, with the specific English Sentiment Model trained off of the standard Sentiment Analysis IMDB dataset[18]. The Flair recurrent neural network (RNN) based model provides cutting edge performance with noted downside of not being specific to the time period or specific textual writing. However, based on the accuracy obtained on IMDB data, which is a standard NLP dataset used throughout the literature[19], we are confident that the added benefit of using the deep neural network outweighs the lack of specificity given by the Loughran-McDonald word count.
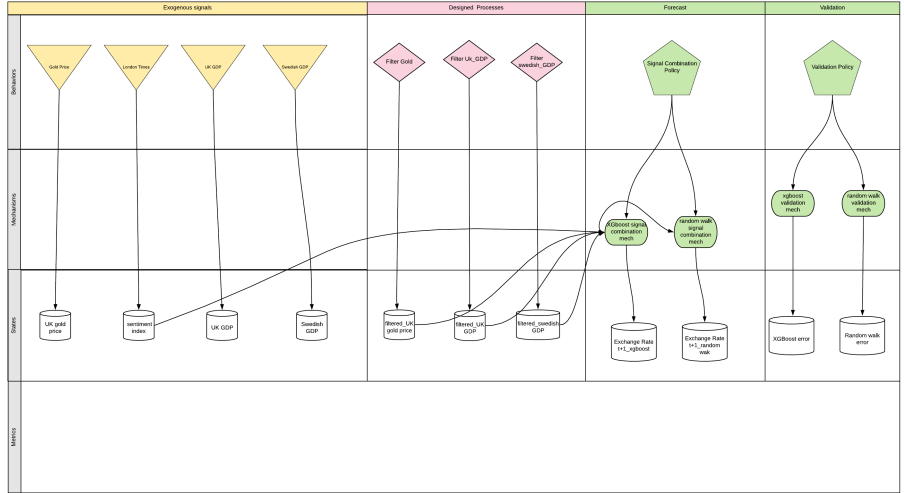
Figure 2: Differential specification of the proposed model

5.2 Complex Dynamical System model

5.2.1 cadCAD

As discussed in the 2.2 subsection, a complex system approach provides a more comprehensive view of the state of the system, can times translate into an increase in prediction capacity.

In cadCAD simulation methodology, there are four layers to a simulation: **Behavior** Policies, Mechanisms, States, and Metrics. Using differential specification using system modeling syntax, it is possible to describe the interaction and information flows between the four layers[20]. **Policies** determine the inputs into the system, and can come from exogenous signals, or from algorithmic policy. **Mechanisms** are functions that take the policy results and update States to reflect input changes. **States** are systems variables that represent quantities at the given point in time. They can either be values that show the change between time periods or 'sink' variables the show the amount of a value at any given point in time. In finance terminology, 'income statement' and 'balance sheet' accounts, respectively. **Metrics** are Key Performance Indicators, KPIs, that are computed from state variables to assess system health

The way to think of cadCAD modeling is analogous to machine learning pipelines which normally consist of multiple steps when training and running a deployed model[21]. There is preprocessing, which includes segregating features between continuous and categorical, transforming or imputing data, and then instantiating, training, and running a machine learning model with specified hyperparameters. cadCAD modeling can be thought of in the same way as states, roughly translating into features, are fed into pipelines that have built-in logic to direct traffic between different mechanisms, such as scaling and imputation.

7

Accuracy scores, ROC, etc are analogous to the metrics that can be configured on a cadCAD model, specifying how well a given model is doing in meeting its objectives. The parameter sweeping capability of cadCAD can be thought of as a grid search, or way to find the optimal hyperparameters for a system by running through alternative scenarios. A/B style testing that cadCAD enables is used in the same way machine learning models are A/B tested, except out of the box, in providing a side by side comparison of multiple different models to compare and contract performance. Utilizing the field of systems identification, dynamical systems models can be used to "online learn" by providing a feedback loop to generative system mechanisms.

The flexibility of cadCAD also enables the embedding of machine learning models into behavior policies or mechanisms for complex systems with an machine learning prediction component.

5.2.2 Kalman Filter

Our cadCAD system takes in the exogenous process variables to the constructed dynamical system and appends them as four separate state variables. Kalman filters[22] are used to mute the volatility changes between years for the UK Gold price, UK GDP and Swedish GDP values. Kalman filters are constructed for each internal state variable to transform the actual value into a filtered value that is more accurate for predictions over time due to the reduced noise.
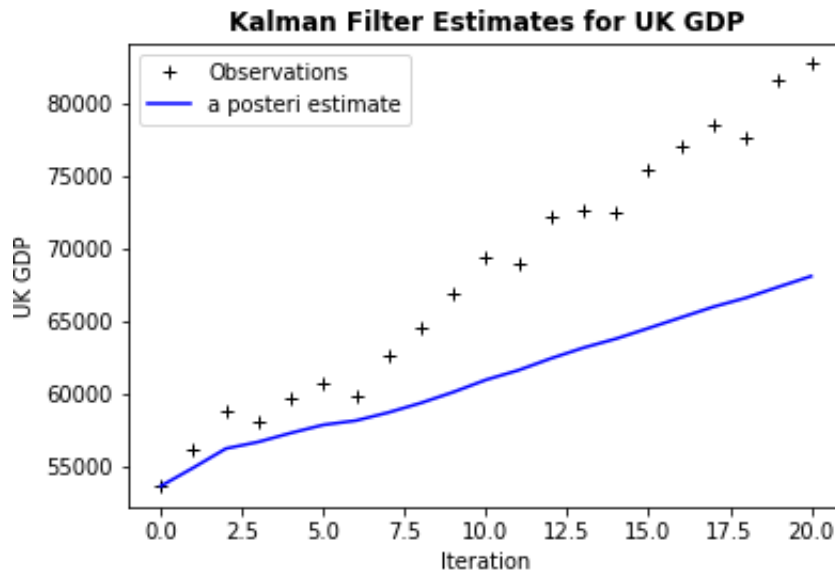

Figure 3: 22 years training of UK GDP Kalman Filter. A slow response to changes in the value results in better forecasting performance

The Kalman filters are trained for 22 years - 1844 to 1864 before the model begins to allow 100 years of model forecasting. The parameters of the Kalman filters are then passed into a prediction function for use in subsequent samplings. As Kalman filters are one step predictors that take the the last observed value and the system parameters to estimate the system state, at each time step, the Kalman filters are retrained. As extremely lightweight algorithms, Kalman filters have been used from Economic Forecasting to Apollo program navigation, and works well in our use case for variable processing[23].

5.2.3   XGBoost

As the main predictor model inside our systems simulation, we use a XGBoost machine learning model[24]. XGBoost is a type of gradient boosting algorithm with the additional attributes of tree penalization, leaf node shrinking, Newton Boosting and randomization[25]. The algorithm which has been very successful on the competition website, Kaggle[25]. Gradient boosting algorithms are, roughly, an ensemble of decision trees that can be optimized over different loss functions when trained on data. The family of Gardient Boosting Models, GBMs, can be applied to both classification, a discrete number of outcomes, such as Pass or Fail, or to regression outcomes, e.x. continuous numerical outputs, like our model.

The XGBoost model constructed within the mechanism of the cadCAD model that combines the Kalman filtered gold, UK GDP, Swedish GDP values as well as the aggregated yearly sentiment. For each timestep of the simulation, the XGBoost model is retrained with the latest Kalman Filter values and the aggregated yearly sentiment to make a one-step prediction. The one-step predicted value is then appended to the prediction state variable that records all values. This iteration continues from years 1856 to 1964, creating a more accurate model overtime and allowing us to see the accuracy of the forecast over time.

5.3   Comparative Model: Random Walk

To validate how well the proposed model is performing, a naive random walk model based off of the following equation was constructed:

$$y(t+1) = y(t)$$

Where t is equal to time and y is the system.

Meese and Rogoff have shown that naive random walk models almost always outperform econometric models, especially in short term horizons.[1] As such, it was theorized that if the proposed model's result is statistically equal to or greater than a random walk, then the model format proposed is sound and warrants more exploration and application.

5.3.1   Order of Events

The various components of the simulation, and its structure, figure 2, have been described above, here we will enumerate the substeps of a system timestep.

```
partial_state_update_blocks = [
    {
        # exogenousProcesses.py
        'policies': {
        },
        'states': {
            'Year': update_timestamp,
            'sentiment':sentiment_exo,
            'uk_gdp': uk_gdp_exo,
            'gold': gold_exo,
            'swedish_gdp':swedish_gdp_exo,
        }
    },
```

Figure 4: Example of a Partial State Update Block

1. In the exogenous processes partial state update block, see figure 4 for an example of the code structure, the sentiment, gdp, and gold data is read into the system and the system year is incremented. See table 1 for a listing of the system state variables.

2. In the designed partial state update block, the model takes the raw inputs of uk gdp, gold, and swedish gdp and updates their respective Kalman filters, and updates the filtered variables by appending the new results.

3. In the forecast partial state update block, the XGBoost model is trained on the filtered values and the corresponding aggregated sentiment, and optimized with squared error; and one timestep is forecasted. The random walk value takes the current year's actual exchange rate value and forecasts next year's value to be the same.

4. In the final partial state update block, validation, the mean squared error is calculated for both the XGBoost forecast and the random walk.

6   Validation

After our simulation was run for 100 years, we plotted the results of the XGBoost model, the random walk model, and compared to the actual values, as seen in figure 5.

State Variables	Purpose
Year	Integer of the simulation year
uk_gdp	UK GDP
sentiment	The NN derived aggregated newspaper sentiment data
gold	Gold value per ounce in pounds
swedish_gdp	Swedish GDP in pounds
uk_gdp_filter	Kalman filtered UK GDP
gold_filter	Kalman filtered Gold value per ounce in pounds
swedish_gdp_filter	Kalman filtered Swedish GDP in pounds
exchange_rate_t+1_xgboost	Predicted krona pound exchange rate from XGBoost
exchange_rate_t+1_random_walk	Predicted krona pound exchange rate from random walk
xgboost_mse	Stepwise MSE from XGBoost
random_walk_mse	Stepwise MSE from the random walk model

Table 1: State variables for the system model

If we look at the the step wise validation of MSE between the random walk and the XGBoost model, we can see that at times the XGBoost outperforms the random walk model, but overall, the naive model performs slightly better, although the difference is negligible. To confirm this, we use the aggregated Mean Squared Error and Root Mean Squared Error over the full simulation, two standard time series evaluation criteria, and recorded in table 2.

Metric	Model	Result
MSE	XGBoost	0.26
RMSE	XGBoost	0.51
MSE	Random Walk	0.25
RMSE	Random Walk	0.5

Table 2: Validation values

Based off of table 2, we can see that XGBoost is on a whole, very slightly outperformed by the random walk model. To test if this is statistically significant or not, we employ the Diebold-Mariano test [26]. The Diebold-Mariano test assumes a null hypothesis of statistical equality between two forecasts. When we run this test off of Python code developed by [27], we receive a test statistic of 0.67 and a pvalue of 0.50, which means that the random walk model and the complex systems XGBoost model introduced in this paper are statistically equivalent, which validates our hypothesis that a complex systems model could equal or outperform a random walk.

If we break the results down further, prior to 1914, the models performed identically, see Table 3. During WWI, our models were nearly identical with a rounding change on the MSEs between models. During the interwar years, the naive model outperforms our model. During WWII, our model outperforms.
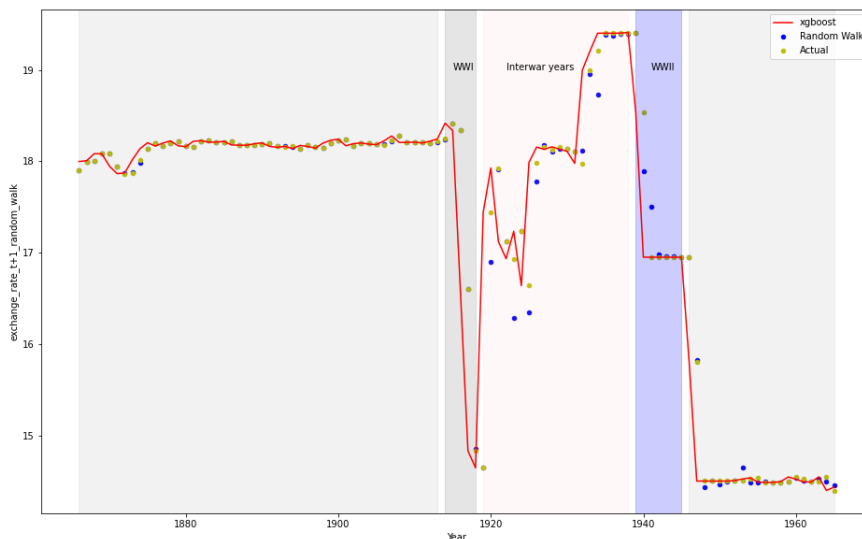
Figure 5: Actual vs predicted. Grey shading means models performed identically, pink the naive model outperformed and blue, our model outperformed

And finally, after WWII, both models were nearly identical, with a rounding error difference on the RMSEs.

7   Conclusion

We downloaded and analysed historical London Times news articles to gather data on how sentiment affected the exchange rate between the Pound and Krona during the century between 1865 and 1965. Through the lens of complex systems theory, we created a simulation model and trained an XGBoost machine learning model to forecast one year exchange rates predictions. The model was retrained every year and during the period studied, our model outperformed a random walk during WWII and had equal performance during all years except the interwar years. This could be an indication that during turbulent times, responsive models can outperform random walks.

As our paper as shown, although ML/DL provides great promise to create more accurate economic models, it is not a panacea. Sometimes simple heuristics, such as the $y(t+1) = y(t)$ random walk can equal or outperform advanced modeling, as our paper shows during some time periods. As big data and ML/DL are becoming more prominent in economics, the importance of quality

Metric	Model	Period	Result
MSE	XGBoost	Pre-WW1	0
RMSE	XGBoost	Pre-WW1	0.05
MSE	Random Walk	Pre-WW1	0
RMSE	Random Walk	Pre-WW1	0.05
MSE	XGBoost	WW1	1.25
RMSE	XGBoost	WW1	1.12
MSE	Random Walk	WW1	1.24
RMSE	Random Walk	WW1	1.12
MSE	XGBoost	Interwar	0.75
RMSE	XGBoost	Interwar	0.86
MSE	Random Walk	Interwar	0.61
RMSE	Random Walk	Interwar	0.78
MSE	XGBoost	WWII	0.28
RMSE	XGBoost	WWII	0.53
MSE	Random Walk	WWII	0.47
RMSE	Random Walk	WWII	0.68
MSE	XGBoost	Postwar	0.09
RMSE	XGBoost	Postwar	0.31
MSE	Random Walk	Postwar	0.09
RMSE	Random Walk	Postwar	0.3

Table 3: Validation values by period

data and parsimony cannot be overlooked. It is often thought that throwing more data at a problem is always the best solution, but quality data and domain expertise in determining feature design and which features to implement cannot be overstated. The difficulty of forecasting prices and ever present battle of correlation vs causality in modeling is not a tool question, but one of approach. As our paper had shown complex modeling and machine learning are incremental lifts to the Economics profession, and can be used successfully to further our forecasting accuracy and understanding of the complex world we live in.

References

[1] Richard A. Meese and Kenneth Rogoff. Empirical exchange rate models of the seventies: do they fit out of sample? *Journal Of International Economics*, 14:3–24, 1983.

[2] Chiranjit Chakraborty and Andreas Joseph. Machine Learning at Central Banks. (ID 3031796), September 2017.

[3] C. Zhang, B. Chen, and J. Pearl. A simultaneous discover-identify approach to causal inference in linear modelsa. In *The Thirty-Fourth AAAI Confer-*

ence on Artificial Intelligence, number 6, Palo Alto, CA, 2020. R-491-L, <http://ftp.cs.ucla.edu/pub/stat_ser/r491-L-reprint.pdf>.

[4] Carlos Cinelli, Daniel Kumor, Bryant Chen, Judea Pearl, and Elias Bareinboim. Sensitivity analysis of linear structural causal models. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1252–1261, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

[5] David Bholat, Stephen Hansen, Pedro Santos, and Cheryl Schonhardt-Bailey. Text mining for central banks.

[6] Federal Reserve Bank of San Francisco, Adam H. Shapiro, Moritz Sudhof, Kanjoya, Daniel Wilson, and Federal Reserve Bank of San Francisco. Measuring News Sentiment. *Federal Reserve Bank of San Francisco, Working Paper Series*, pages 01–22, January 2017.

[7] cadCAD – A Python package for designing, testing and validating complex systems through simulation.

[8] Shermin Voshmgir and M. Zargham. Foundations of cryptoeconomic systems. 2019.

[9] Sergio Focardi. Is economics an empirical science? if not, can it become one? *Frontiers in Applied Mathematics and Statistics*, 1:7, 2015.

[10] Gale - the Times Digital Archive, 1785-2013. https://www.gale.com/c/the-times-digital-archive.

[11] Håkan Lobell. *Historical Monetary and Financial Statistics for Sweden: Exchange rates, Prices and Wages, 1277–2008*, volume I, chapter 6 - Foreign exchange rates 1804–1914, pages 291–339. Svergies Riksbank, 2010.

[12] Jan Bohlin. *Historical Monetary and Financial Statistics for Sweden: Exchange rates, Prices and Wages, 1277–2008*, volume I, chapter 7 - From appreciation to depreciation-the exchange rate of the Swedish krona, 1913-2008, pages 340–411. Svergies Riksbank, 2010.

[13] Samuel H. Officer, Lawrence H.; Williamson. The price of gold, 1257-present, 2019. http://www.measuringworth.com/gold/.

[14] Rodney Edvinsson. *House Prices, Stock Returns, National Accounts, and the Riksbank Balance Sheet, 1620–2012*, volume II, chapter 4 - The Gross Domestic Product of Sweden within present borders, 1620–2012, pages 101–182. Svergies Riksbank, 2014.

[15] Bank of england, real gross domestic product at market prices in the united kingdom. FRED, Federal Reserve Bank of St. Louis.

[16] Tim Loughran and Bill McDonald. When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks. SSRN Scholarly Paper ID 1331573, Social Science Research Network, Rochester, NY, March 2010.

[17] Alan Akbik, Duncan Blythe, and Roland Vollgraf. Contextual string embeddings for sequence labeling. In *COLING 2018, 27th International Conference on Computational Linguistics*, pages 1638–1649, 2018.

[18] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.

[19] Papers with Code - IMDb Benchmark (Sentiment Analysis).

[20] cadCAD Differential Specification Syntax Key, September 2019.

[21] 6.1. Pipelines and composite estimators — scikit-learn 0.23.2 documentation.

[22] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, March 1960.

[23] Wolfgang Schneider. Analytical uses of Kalman filtering in econometrics — A survey. *Statistical Papers*, 29(1):3–33, December 1988.

[24] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.

[25] Synced. Tree Boosting With XGBoost – Why Does XGBoost Win "Every" Machine Learning Competition?, October 2017.

[26] Francis Diebold and Roberto Mariano. Comparing predictive accuracy. *Journal of Business Economic Statistics*, 13(3):253–63, 1995.

[27] John Tsang. johntwk/Diebold-Mariano-Test, October 2020. original-date: 2017-12-07T15:33:38Z.