

# Modern Advanced Analytics Platforms and Predictive Models for Stock Price Forecasting: IBM Watson Analytics Case

Sergey Yablonsky

The Graduate School of Management of St. Petersburg University, St. Petersburg, Russian Federation  
yablonsky@gsom.spbpu.ru; yablonsky.serge@gmail.com

Ilias Faizullov,

The Graduate School of Management of St. Petersburg University, St. Petersburg, Russian Federation  
st024827@student.spbu.ru

## Abstract

*The primary purpose of this paper was to provide an in-depth analysis of the ability of modern analytical platforms (using IBM Watson Analytics as an example) to generate predictive models for stock prices forecasting in comparison with traditional analytical econometric platforms and models. Series of stock predictive models based on the suggestions of IBM Watson Analytics have demonstrated results, which are superior to all other models. In terms of forecasting accuracy, they beat all models except for the Random Walk. The simulation has demonstrated high returns for most of the suggested models.*

## 1. Introduction

In 2011 report by the McKinsey Global Institute (MGI) it was mentioned that only the United States needs 1.5 million more data-literate managers to meet the demands of the data-driven enterprise [1]. Furthermore, digital flows now exert a larger impact on GDP growth than the centuries-old trade in goods, according to a new 2016 MGI report [2]. That's why effective business analytics is a focus for business leaders across the globe in increasing numbers [3].

Advanced analytics platforms have always been tools built for statisticians and data scientists. But in contrast to previous years Gartner noted in [4] that these tools are "increasingly being directed at business analysts and 'citizen data scientists.'" Gartner defines a "citizen data scientist" as a "person who creates models that leverage predictive or prescriptive analytics, but whose primary job function is outside of the field of statistics and analytics. This person is not typically a member of an analytics team (for example, the analytics center of excellence) and does not necessarily have a job description that lists analytics as their primary role. This person is typically in a line of business that is outside of IT and outside of a BI team." [4]. Citizen data scientists are "power users" who will be able to perform simple and moderately sophisticated analytic tasks that would previously have required more

expertise. New tools will also make highly skilled data scientists more productive, enabling them to churn out more analysis in the same amount of time. Thus, Gartner's rankings on its well-known "completeness of vision" and "ability to execute" axes now ride as much on "ease of use" as they do on offering numerous and novel analytical algorithms.

IBM Watson Analytics that appeared in 2015 is among the new modern advanced analytics platforms easy to use and directed at business analysts and citizen data scientists. IBM hopes Watson Analytics has the potential to give business users access to a new user experience for data discovery, featuring an integrated analytic workflow that includes self-service data preparation, natural language query generation and exploration, and automatic pattern detection and prediction — to allow business analysts and citizen data scientists to find valuable insights in data without advanced skills" [5].

Goal of this research is to assess the ability of IBM Watson Analytics to suggest effective predictive models for stock price forecasting.

Research question of this work is as follows: does modern advanced analytical platforms (using IBM Watson Analytics as a case) suggest effective predictive models for stock forecasting, in comparison with traditional analytical econometric platforms and models?

Research objectives of this paper are:

- To construct and evaluate theoretically based econometric models for stock prices forecasting.
- To construct econometric models for stock price forecasting using factors, suggested by IBM Watson Analytics prediction function.
- To compare the performance of theoretically based, and IBM Watson Analytics suggested models.

In a course of this research, we will evaluate the abilities of mentioned analytical platform to serve the needs of traders. In case of traditional analytics platforms (statistical packages), we will use theoretically based econometric models, and in case of IBM Watson analytics we will let the platform to suggest optimal models by itself. This approach has a

potential problem: lack of the theoretical groundings. For a trader, it may appear be irrelevant, since he/she mostly cares about the accuracy of forecasts, however, without theoretical basis it is impossible to guarantee the stability of the model: it could have just happened that the factors, which affected the predicted variables, are spuriously correlated.

In the research, we will build series of models. The first one will be standard random walk models for currency's exchange rates. It will be used for a comparison with other models, since they will make any sense only in case if they outperform the random walk. Another series of predictive models for currency's exchange rates will be constructed using simple one-factor model that use price of the most exported commodity as a predictor. The dynamic of stock market will be analyzed by applying Capital Asset Pricing Model to the blue chips of the United States stock exchange: Microsoft, Apple, IBM, Bank of America, Walmart, and P&G. The US stock market was chosen because of the necessity of operating under the Effective Market Hypothesis, which more likely to be true in the developed market, rather than the emerging one.

Final series of predictive models will be constructed in Gretl – an open-source statistical package, but in this case, factors will be chosen based on the suggestions of IBM Watson Analytics, which automatically determines drivers of a given variable. Predictive accuracy of the forecasts generated by aforementioned models will be estimated by two characteristics: Mean Absolute Percentage Errors, and potential profitability. The latter characteristic will be assessed through the results of a trading simulation experiment during which we will imitate real-life trading using all of the models we have constructed.

Simulation will be run in accordance with rules as follows:

- If model predicts, that price of the asset will rise in the next period, an investor makes a decision to buy the asset.
- If model predicts, that price of the asset will fall in the next period, an investor makes a decision to sell the asset.
- If an investor bought the asset, he would sell it in the next period regardless of its new price.
- If an investor sold the asset, he would buy it back in the next period, regardless of its new price.

At the end of the prechosen period, investors stops and calculates his/her returns, which will be used as an indicator of forecasting accuracy of the model. In order to have more reliable indicator of the forecasting accuracy, we will run a model, simulating real life trading. Rules of the model are simple, if it

anticipates, that asset's price will increase in the next period, than an investor takes the decision to buy the asset, with the intention to sell it afterwards. Depending of the actual change of the prices, such operations could bring profits or loses.

## **2. Different types of advanced analytics**

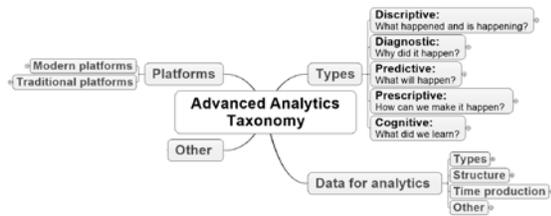
### **2.1. Advanced analytics taxonomy**

What exactly is “analytics”? Davenport and Harris [6] define analytics as “the ability to collect, analyze, and act on data”. The field of analytics is broken down into three categories: descriptive, predictive, and prescriptive analytics [7-10]. Today some of researchers and providers like IBM add cognitive or smart analytics [8, 9]. How do these four categories help to define advanced analytics? In our analysis we argue that rather than searching for a single acceptable definition, a better approach would be to develop a classification system or taxonomy [12]. A clear and precise description and structuring of the information in the advanced analytics domain are prerequisites for a common research. Taxonomies and other types of controlled vocabularies are the preferred means to achieve such a common understanding by specifying the terms of the domain, disambiguating them from each other, controlling synonyms, and structuring the domain via term relationships. For conceptual grounding of the categories in advanced analytics taxonomy we use definitions from different information resources [3-11]. The pilot version of the advanced analytics taxonomy is shown on Figure 1(hierarchy of 2 first layers of taxonomy concepts). Taxonomy includes a set of basic concepts, a set of relations holding between those concepts, and a set of instances – international and local analytic platform service providers. The total number of all taxonomy features (>100) is too complex to be represented here in its entirety, but an example of taxonomy is provided in order to demonstrate both the process of classification and the intermediate result. We suggest that although classification systems have been used in the business and management disciplines, the more advanced quantitative methodologies have not yet been widely used. The future research could usefully build on these techniques to construct enhanced classification systems of advanced analytics approaches across a variety of dimensions in addition to the basic concepts of type, data, and platform (Figure 1).

### **2.2. Modern analytics platforms**

According to Gartner's February 2016 Magic Quadrant [7], Gartner introduced its new definition of modern analytics platform, which suggests that

analytics platforms are in the last stages of moving from being IT-led to being business-led.



**Figure 1.** Advance analytics taxonomy. Hierarchy of 2 first layers of main concepts.

A modern BI and analytics platform supports IT-enabled analytic content development. It is defined by a self-contained architecture that enables nontechnical users to autonomously execute full-spectrum analytic workflows from data access, ingestion and preparation to interactive analysis, and the collaborative sharing of insights [7].

By contrast, traditional BI and analytics platforms are designed to support modular development of IT-produced analytic content, specialized tools and skills, and significant upfront data modeling, coupled with a predefined metadata layer, is required to access their analytic capabilities.

The most significant difference between a modern analytics platform and a traditional platform is stated in Table 1 [4,5]. Gartner says that as analytics has become increasingly strategic to most businesses and central to most business roles, every business is an analytics business, every business process is an analytics process and every person is an analytics user.

**Table 1.** High-Level Comparison of Traditional and Modern BI and Analytics Platforms

Analytics Workflow Component	IT-Centric Reporting and Analysis Platform	Modern BI and Analytics Platform
Data Source	Upfront dimensional modeling required (IT-built star schemas)	Upfront modeling not required (flat files/flat tables)
Data Ingestion and Preparation	IT-produced	<b>IT-enabled</b>
Content Authoring	Primarily IT staff, but also some power users	<b>Business users</b>
Analysis	Structured ad hoc reporting and analysis based on a predefined model	<b>Free-form exploration</b>

"It is no longer possible for chief marketing officers (CMOs) to be experts only in branding and

ad placement," states Gartner. "They must also be customer analytics experts. The same is true for the chief HR, supply chain and financial roles in most industries" [13]. In [4, 9] IBM Watson Analytics is positioned as a modern analytics platform.

### 3. Evaluation of the forecasting accuracy of IBM Watson Analytics

#### 3.1. Data description

In the Table 2, we can see a description of the data we used in the stock market forecasting experiments. Variables are classified into four categories: stock prices, prices of resources (gold, oil, and natural gas), values of the market indexes, and currency's exchange rates. Observations cover the period from 01.30.2015 to 01.04.2016

We used two types of software to run the predictive modeling: Gretl statistical package and IBM Watson Analytics. Type of models, which is marked as IBM+Gretl in the Table 2, was built in a steps as follows: after uploading the dataset to IBM Watson Analytics, the predictive function was applied. It has suggested predicting factors for each target variable (stocks and currency's exchange rates), after that, simple two-factor regression models were built in the SPSS, using suggested by the IBM Watson Analytics predictive factors as independent variables. The random walk models are basically just ARIMA (0,1,0) models. They will be used just as a basis for comparison.

**Table 2.** Data description (Source: Finam [16])

Software	Model	Variables	Number of observations
Gretl	Random Walk Models	ERO/USD	429
		USD/CAD	426
		USD/YEN	428
		USD/ZAR	426
		USD/NOR	425
		USD/CNY	393
	One-Factor models	USD/RUB	426
		USD/NOR	363
		USD/ZAR	
		USD/RUB	
	BRENT		
	CAPM	Gold	286
		S&P 500	
		BAC	
IBM			
MSFT			
P&G			
IBM+Gretl	Two-Factor	Walmart	225
		Apple	
		S&P 500	
		DJI	

models	RTS
	Nikkei
	CSI
	FTSE
	Shanghai
	NASDAQ
	Gold
	Natural Gas
	Brent
	ERO/USD
	USD/CAD
	USD/YEN
	USD/ZAR
	USD/NOR
	USD/CNY
	USD/RUB
	Exxon Mobil
	Chevron
	BAC
	IBM
P&G	
Walmart	
Apple	

One-factor models are predicting the currency's exchange rates based on the prices of the most exported commodities (oil, and gold). CAPMs predict the prices of the stock. It was build using week prices of the "blue chips" of the US stock market. Role of the average market indicator was played by the S&P 500 index. Interest rate of the 4 week treasure bills was used as the risk free rate (Rfr=2%). Return on assets is calculated as the difference between stock's price in moment t and stock's price in moment t-1, divided by the stock price in the moment t-1:

$$R=(P_t-P_{(t-1)})/P_{(t-1)} \quad (1)$$

### 3.2. Forecasting stock prices with theoretically based models.

**3.2.1. Results of the Random walk models for currencies.** Random walk model is the basis for comparison for any other forecasting model, as any predictive model makes sense only if it beats the random walk. Using Gretl statistical and econometrical package, we have built ARIMA(0,1,0) time series models, which are equivalent to the simple Random walk. In the Table 3, we can see the error metrics for the random walk model for currencies exchange rates.

As we can see in the Table 3, random walk models have produced quite small mean percentage errors and mean absolute percentage errors, with the exception for USD to South African Rand Exchange rate (ZAR). It might give an impression that random walk performs greatly, however, as it is supported by

Elliot, G (2013), for the purposes of profiting from the differences in the exchange rates, it's more important to foresee the direction of change, rather than to give more accurate estimation. Low percentage error in Random Walk case could be caused by the fact that the forecasted value differs from the previous observation only by small random value.

**Table 3.** Percentage errors of Random Walk models

Model	MPE	MAPE
EUR/USD	0.004209	0.004209
USD/CAD	-0.0011137	0.056509
USD/NOK	-0.077933	0,077933
USD/RUB	-0.27649	0,27649
USD/ZAR	8.1111	8.1111
USD/CNY	-0.0089793	0.0089793
USD/JPY	-0.032764	0.032764

Our next step is to estimate potential profitability of trading main currencies using random walk model. For that purpose, we have run the simulation test in Excel 2013, using rules as follows: if the investor expects appreciation of the asset, then he buys it, and vice versa. The results are shown in the Table 4. We have used 30 last forecasted values of each currency's exchange rates, for an imitation of real life trading.

**Table 4.** Results of the simulation of Random Walk

Model	Profitability
EUR/USD	0,39%
USD/CAD	-6,20%
USD/NOK	-1,38%
USD/RUB	-17,45%
USD/ZAR	-0,46%
USD/CNY	0,96%
USD/JPY	1,64%

As expected, results of the simulation reveal that Random walk model is absolutely unfit for trading, in 4 out of 7 cases, the profitability is negative, especially in case of Ruble, which has shown over -17% loses. Even positive examples have very low profitability. The average return is -3.2%, and if it were real life trading, than the loses would be even bigger, as there are transaction costs and time lags. Thus, it is safe to conclude that random walk model is completely unfit for real life application. As we can see in the Table 2, random walk models have produced quite small mean percentage errors and mean absolute percentage errors, with the exception for USD to South African Rand Exchange rate (ZAR). It might give an impression that random walk

performs greatly, however, as it is supported by Elliot, G (2013), for the purposes of profiting from the differences in the exchange rates, it's more important to foresee the direction of change, rather than to give more accurate estimation. Low percentage error in Random Walk case could be caused by the fact that the forecasted value differs from the previous observation only by small random value.

**3.2.2 Currency's exchange rates forecasting using factor models.** Table 5 presents the description of factor models. In accordance with Dominico, F. (2015) we have built predictive model for currency's exchange rate forecasting using prices of mostly traded commodities as predictors. Models were built in Gretl econometrical package using "ordinary least square" option.

**Table 5.** Description of factor models for currencies.

Model	Model's Parameters		Model's Statistic	
	Coefficient	Sig.	R-squared	MAPE (%)
USD/NOK			0,7954	3,6611
const	10,1061	<0,0001		
Brent	-0,03916	<0,0001		
USD/RUB			0,7708	4,0127
const	83,0842	<0,0001		
Brent	-0,48308	<0,0001		
USD/ZAR			0,1699	14,503
const	26,4943	<0,0001		
Gold	-0,01142	0,0003		
USD/CAD			0,7716	4,19840
const	1,4658	<0,0001		
Brent	-0,00414812	<0,0001		

All factors are statistically significant and they have expected influence on every currency (the higher the price of the commodity, the lower USD exchange rate). However, these models demonstrate bigger mean percentage errors than the random walk. In that sense, they don't beat the random walk.

Three out of four models have high R-squared (>0.7), which implies good explanatory power of models. The only exception is USD/ZAR model, which has very low R-squared (=0.169) and the highest Mean Absolute Percentage Error (14%). This result leads us to the thoughts that, gold isn't the main export product in South Africa anymore.

Our next step is to estimate potential profitability of trading main currencies using simple one factor regression. For that purpose, we have run the

simulation test in Excel 2013. We have used 30 last forecasted values of each currency's exchange rates, for an imitation of real life trading.

As it is shown in the Table 6, trading with factor models brings way higher returns, that just random walk, because factor models manage to generate more accurate predictions of the direction of price's change. Average return for this model is 26%.

**3.2.3 Stock forecasting using CAPM model.** Using "ordinary least square" function in Gretl statistical package, we have built CAPM for every of stocks as follows: Apple, IBM, Microsoft, Procter & Gamble, Walmart, and Bank of America. As a factor we have used the risk premium:

$$RP = (\mu_M - R_0), (2)$$

where  $\mu_M$  is return on S&P index, and  $R_0$  is four weeks treasure bill interest rate.

**Table 6.** Results of the simulation of the factor models

Model	Factor	Profitability
USD/CAD	Brent	21,51%
USD/RUB	Brent	31,32%
USD/ZAR	Gold	22,84%
USD/NOK	Brent	28,75%

As we can see it in the Table 7, CAPM model produce quite poor results both in terms of explanatory power (low R-squared) and accuracy of forecasts, sometimes mean percentage errors exceed 100% (Average MAPE = 177%), meaning that the forecasts is radically different with the reality. Despite the fact that in all cases, risk premium as a factor was significant, and R-squared is tolerable (except for Walmart case), the models appear to be unfit for the actual forecasting. Because of huge deviations of forecasted values from the actual ones.

**Table 7.** Description of CAPM for stocks.

Model	Model Parameters		Model's Statistic	
	Coefficient	Sig.	R-squared	MAPE (%)
Bank Of America			0.5493	574.13
const	0.000181202	<0.0001		
RP	1.32574	<0.0001		
Microsoft			0.4642	258.91
const	0.00130152	0.0833		
RP	1.13318	<0.0001		
Walmart			0.2148	59.562
const	-0.0015747	0.0246		
RP	0.592545	<0.0001		
IBM			0.4535	72.152
const	-0.000336033	<0.0001		
RP	0.919689	<0.0001		

P&G			0.4593	66.181
const	-0.000790775	0.0787		
RP	0.671332	<0.0001		

Our next step is to estimate potential profitability of trading blue chips stocks using CAPM model. For that purpose, we have run the simulation test in Excel 2013. We have used 30 last forecasted values of each currency's exchange rates, for an imitation of real life trading. Results of the trading simulation (Table 8) confirms the point that CAPM is unfit for stock market forecasting. CAPM has generated significant potential outcome only in 2 out of 6 cases, in two cases, the results were negative, and the last two have demonstrated negligible profits, which would not even cover transactional costs. Average return is 5%, which demonstrates that despite huge deviations of forecasted values from actual ones, in some cases CAPM still correctly predict the direction of change.

**Table 8.** Results of the CAPM simulation

Model	Profitability
BAC	17,74171%
IBM	19,201815%
MSFT	3,278492%
P&G	-3,42963%
Walmart	1,295297%
Apple	-8,38422%

### 3.3 Forecasting stock market using IBM Watson analytics.

**3.3.1 Models for stock forecasting.** We have used free version of IBM Watson analytics to conduct our experiment. After uploading our dataset consisting of 26 variables, IBM Watson Predict option has automatically processed and analyzed uploaded data. The result is a set of suggested predictive factors that drive any given variable. Based on the predictive power of the model, estimated by Watson Analytics, we have chosen the most promising ones. Forecasting of stock prices and currencies exchange rates using IBM Watson will be done using IBM Watson analytics "Predict" function in two steps:

1. Choosing factors, which IBM Watson Analytics Suggest as the best predictors
2. Building two factor regression using Ordinary Least Square method in Gretl statistical package

Using suggested drivers of predicted values, we have built regression models in Gretl statistical package for each of the observed currency's exchange rate. The results are presented in the Table 9.

Analyzing the results, we can see that three out of thirteen models (IBM 1, IBM 2, and Chevron 2) turned out to be statistically insignificant. That strange result could be explained by the fact that some potentially important predictors were not included in the uploaded dataset. IBM Watson just didn't have enough data to generate good models for these stocks.

**Table 9.** Description of models built based on IBM Watson Analytics platform

Model	Model Parameters		Model's Statistic	
	Coefficient	Sig.	R-square	MAPE(%)
Exxon Mobil 1			0.7129	2.347
const	-4.572	0.0226		
Gold	0.007250	<0.0001		
Futsee	0.0339528	<0.0001		
Exxon Mobil 2			0.82757	1.8791
const	-46.8102	<0.0001		
DJI	0.0405594	<0.0001		
Gold	0.00467673	<0.0001		
Exxon Mobil 3			0.76269	2.2036
const	-53.5047	<0.0001		
SP 500	0.0398747	<0.0001		
Gold	0.0463417	<0.0001		
IBM 1			0.87452	11.2
const	78.6593	<0.0001		
Brent	1.19685	<0.0001		
NKK225	0.00063672 2	0.0640		
IBM 2			0.8174	7,54
const	241.98	<0.0001		
NASDAQ100	0.00448579	0.0916		
USDZAR	-8.25771	<0.0001		
P&G			0.3058	3.19
const	151.348	<0.0001		
USDJPY	-0.683958	<0.0001		
Brent	0.197202	<0.0001		
Bank of America			0.8083	3.49
const	-19.479	<0.0001		
Natural Gas	2.11465	<0.0001		
NASDAQ100	0.00683298	<0.0001		
Apple 1			0.9280	2.02
const	390.097	<0.0001		

DJI	0.00533233	<0.0001		
USDCNY	-57.7748	<0.0001		
Apple 2			0.9347	1,20
const	-9.95716	0.0590		
Brent	0.879771	<0.0001		
NASDAQ100	0.0186525	<0.0001		
Walmart 1			0.6272	1,50
const	191.434	<0.0001		
USDZAR	-5.50512	<0.0001		
NKK225	-0.0025033 4	<0.0001		
Walmart 2			0.7842	1,49
const	-33.1095	<0.0001		
NKK225	-0.0030498 3	<0.0001		
Footse100	0.0248156	<0.0001		
Walmart 3			0.6013	1,73
const	151.821	<0.0001		
Brent	0.695071	<0.0001		
USDJPY	-0.971299	<0.0001		
Chevron 1			0.7544	4,29
const	-76.5478	<0.0001		
Gold	0.0393618	<0.0001		
Footse100	0.0192772	<0.0001		
const	119.11	<0.0001		
NASDAQ100	0.0063668	0.0502		
USDZAR	-3.97567	<0.0001		
Coke 1			0.6073	2,03
const	43.2875	<0.0001		
Gold	0.00589016	<0.0001		
Natural Gas	-3.42148	<0.0001		
Coke 2			0.4070	7,54
const	-18.4188	0.0002		
USDCNY	7.84844	<0.0001		
Gold	0.00890342	<0.0001		

R- squared is high or at least tolerable in all cases with the exception for P&G. Additionally, there are two models with a borderline explanatory power – Coke 2 and Chevron 2, R-squared equals 0.407 and 0.425 respectively. Mean percentage errors are quite low, but still they are higher than that of a random walk model.

As a next step, we have estimated potential profitability of trading stocks using regression models, with factors suggested by IBM Watson analytics. For that purpose, we have run the simulation test in Excel 2013. The results are shown in the Table 10. We have used 30 last forecasted values of each currency's exchange rates, for an imitation of real life trading.

**Table 10.** Results of the simulation of IBM predictive models.

Model	Factors	Profitability
Apple 1	DJI	45,29%
	USDCNY	
Apple 2	Brent	42,63%
	Nasdaq 100	
Exxon Mobil 1	Gold	9,92%
	Futsee 100	
Exxon Mobil 2	Gold	46,89%
	DJI	
Exxon Mobil 3	SP 500	44,79%
	Gold	
IBM 1	Brent	16,94%
	NKK225	
IBM 2	NASDAQ100	18,16%
	USDZAR	
PG	USD/JPY	-6,11%
	Brent	
Bank of America	Natural Gas	37,98%
	NASDAQ100	
Chevron 1	Futsee 100	19,07%
	Gold	
Chevron 2	NASDAQ100	53,17%
	USDZAR	
Walmart 1	USDZAR	2,00%
	NKK 225	
Walmart 2	NKK 225	23,98%
	Futsee 100	
Walmart 3	Brent	-2,70%
	USD/JPY	
Coke 1	Gold	0,033%
	Natural Gas	
Coke 2	Gold	-16,52%
	USD/CNY	

We have ambivalent results, on one hand; some of the models have demonstrated superior results

during the simulation (Apple 1, Apple 2, Exxon Mobil 2, Exxon Mobil 3, and Bank of America), but on the other hand, three models have demonstrated negative result (P&G, Walmart 3, and Coke 2), and one has shown negligibly small profitability (Walmart 1). The lowest results were demonstrated by those models, which turned out to be insignificant (IBM 1, IBM 2). As it was mentioned before, the reason for these results could be absence of some important factors in the dataset.

Overall, IBM Watson generated models have shown results that exceed any other in terms of potential profitability. Average return is 20%, which is way better than that of CAPM. However, there is a problem of separating profitable models from unprofitable ones, and the stability of the desirable performance over the time is still in question.

**3.3.2 Models for currency's exchange rate forecasting.** We have used IBM Watson analytics to conduct our experiment. After uploading our dataset consisting of 26 variables, IBM Watson Predict option has automatically processed and analyzed uploaded data. The result is a set of suggested predictive factors that drive any given variable. Based on the predictive power of the model, estimated by Watson Analytics, we have chosen the most promising ones. Forecasting of stock prices and currencies exchange rates using IBM Watson will be done using IBM Watson analytics "Predict" function in two steps:

1. Choosing factors, which IBM Watson Analytics Suggest as the best predictors
2. Building two factor regression using Ordinary Least Square method in Gretl statistical package

Using suggested drivers of predicted values, we have built regression models in Gretl statistical package for each of the observed currency's exchange rate. The results are presented in the Table 11.

**Table 11.** Description of currency's exchange rate models.

Model	Model Parameters		Model's Statistic	
	Coefficient	Sig.	R-squared	MAPE (%)
EUR/USD1			0,409	2,8818
const	1,02361	<0,0001		
Gold	0,0002999	<0,0001		
PG	-0,003398	<0,0001		
USD/CNY1			0,898	0,0547
const	6,94919	<0,0001		
Brent	-0,011016	<0,0001		

Shanghai	-1,60684e-05	0,0335		
USD/CNY2			0,919	0,4768
const	7,22139	<0,0001		
Brent	-0,00990401	<0,0001		
NKK225	-2,03678e-05	<0,0001		
USD/JPY1			0,824	3,1761
const	103,254	<0,0001		
BankAmerica	1,73662	<0,0001		
Gold	-0,009451	<0,0001		
USD/JPY2			0,864	1,8067
const	109,078	<0,0001		
Gold	-0,020616	<0,0001		
NKK225	0,0018470	<0,0001		
USD/NOK1			0,840	0,9656
const	14,2993	<0,0001		
Gold	-0,003589	<0,0001		
Natural Gas	-0,776106	<0,0001		
USD/NOK2			0,819	3,5709
const	9,92649	<0,0001		
Natural Gas	-0,033129	0,0515		
Brent	-0,032328	<0,0001		
USD/ZAR			0,907	3,6134
const	21,3444	<0,0001		
Natural Gas	-0,977925	<0,0001		
Brent	-0,108871	<0,0001		
USD/RUB			0,9182	5,8449
const	104,251	<0,0001		
Brent	-0,564791	<0,0001		
Shanghai	-0,00349858	<0,0001		

As we can see it in the Table 11, all of the models are statistically significant and have high values of R – squared, with the exception for the Euro to USD exchange rate model. In terms of percentage errors, models still are not capable of beating the Random Walk.

As a next step, we have estimated potential profitability of trading currencies using regression models, with factors suggested by IBM Watson analytics. For that purpose, we have run the simulation test in Excel 2013. We have used 30 last forecasted values of each currency's exchange rates, for an imitation of real life trading. Results of the simulation tests are shown in the Table 12. In all

cases except for Euro to USD, models were able to produce positive results, but the profitability is much lower than that of the stock predicting models (10% vs. 26%), this result is quite surprising. It once again raises question of stability of performance of econometrical models.

**Table 12.** Results of the simulation for currencies

Model	Factors	Profitability
EUR/USD	Gold	0,06876367
	PG	
USD/CNY 1	Brent	-0,0136448
	Shanghai	
USD/CNY 2	Brent	0,00554467
	NKK225	
USD/JPY 1	BankAmerica	0,13238344
	Gold	
USD/JPY 2	Gold	0,12007505
	NKK225	
USD/NOK 1	Gold	0,0160639
	Natural Gas	
USD/NOK 2	Natural Gas	0,08606294
	Brent	
USD/ZAR	Natural Gas	0,12584994
	Brent	
USD/RUB	Brent	0,39969893
	Shanghai	

## 4. Conclusion

We have built a series of predictive models for stock price forecasting and currency exchange rate forecasting. First series were based on the Random Walk model. It was chosen as basis for comparison with other models, as it necessary for any predictive model to outperform random model in order to make at least some sense.

Random walk models have shown unbeatably small deviations of forecasted values from the actual ones, but the random walk model fails to correctly predict the direction of change, therefore it is completely unfit for the purposes of trading. Another type of currency's exchange rate forecasting model we employed is one factor regression, which uses price of the most exported commodity as a predictor. In terms of deviations of forecasts from actual values, they failed to beat the Random Walk, but in terms of potential profitability, as it was demonstrated by the simulation, they easily outperformed the Random Walk, by demonstrating returns on the level of 20-30%.

Next models we built are CAPM models for the "blue chips" with the index S&P 500 as an average

market asset. CAPM has shown poor results in terms of both forecasting accuracy and potential profitability. Its deviation from actual values sometimes exceeded 100%, and only one model has shown substantial returns during the simulation.

Series of stock predictive models based on the suggestions of IBM Watson Analytics have demonstrated results, which are superior to all other models. In terms of forecasting accuracy, they beat all models except for the Random Walk. Additionally, the simulation has demonstrated high returns for most of the suggested models, with the exception for four models with negative and unsubstantial returns. Results of currency's exchange rate forecasting using IBM Watson were worse than that of a simple one-factor regression models, it still beats the Random Walk in potential profitability. It raises the question of spurious correlation between the variables.

Overall, IBM Watson Analytics is capable of suggesting effective predictive models. However, it doesn't provide users with detailed description of the nature of the interdependencies between the variables. It requires further analysis in order to compute actual forecast of the variables in question.

The results has shown that in terms of deviations of forecasts from the actual values of observed variables (measured in terms of Mean Absolute Percentage Errors), the Random Walk is unbeatable. However, when it comes to the potential profitability of the models (assessed trough trading simulation), theoretically based models has shown worse results, that IBM Watson Analytics suggested models, with the exception of the models, based on the prices of most exported commodities. This result could be explained by the fact, that IBM Watson Analytics didn't specify the nature of interdependencies between the variables, meaning that further analysis is required in order to determine the exact econometric equation. Overall, the effectiveness of IBM Watson Analytics as an effective tool for predictive models suggestion was confirmed. The research indicated that IBM Watson Analytics platform suggests effective predictive models, however, further analysis is required in order to build the most effective predictive model.

### 4.1. Theoretical implications

Using this methodology, similar researches of other analytical platforms could be conducted.

The research can serve as a base for further studies of how big data challenges in financial sector could be tackled using analytical platforms.

The inability of CAPM to adequately predict stock prices even on the developed stock market was

confirmed, therefore the Effective Market Hypothesis is not met on the US stock market.

The research has both confirmed and questioned the unbeatable random walk: in terms of the deviation measures, the random walk remains unbeatable, but from the perspective of forecasting the direction of change, it is outperformed by both theoretically based models, and by those that were suggested by IBM Watson Analytics.

#### 4.2. Managerial implications

The research provides individual traders with tight budget constraints with the no costs combination of analytical platforms (IBM Watson Analytics as a guide, and Statistical Package (Gretl) for the construction of the final model). This combination could prove to be quite effective, since IBM Watson Analytics is the only tool which is capable of suggesting predictive models without preliminary theoretical work.

The study has identified the analytical functions, which analytical platform should be able to perform in order to address the business tasks of the contemporary financial organizations.

#### 5. Limitations

Ability of IBM Watson Analytics to suggest predictive models was compared only with mostly common used econometric models.

There are some collateral theoretical results: the theory that currency's exchange rate could be effectively predicted using the price of the mostly exported commodities was confirmed, however, is models have limited applicability, since they could predict exchange rates only for those currencies which are strongly connected to one particular commodity. In other words, it applies only to resource exporting economies.

This study was conducted with the use of open source data gathered from the Finam website [16]. Access to the more possible variables harness the possibility for Watson Analytics to generate better predictive models.

Finally, our simulations were run under the assumption that an investor has instant access to all information, needed for the model building, and that an investor can strike deals instantly, before the market reacts on the changes. Simulation of the potential profitability is made under the assumption that an investor have access to all necessary information and reacts on it instantly.

#### 6. References

[1] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. Big data: The next frontier for innovation,

competition, and productivity, McKinsey Global Institute Report, May 2011.

[2] Manyika, J., Lund, S., Bughin, J., Woetzel, J., Stamenov, K., and Dhingra, D. Digital globalization: The new era of global flows, McKinsey Global Institute Report, February 2016.

[3] Isson, J. and Harriot, J. Win with Advanced Business Analytics: Creating Business Value from Your Data. John Wiley & Sons, 2013, 398 p.

[4] Kart, L., Herschel, G., Linden, A., Hare, J. "Magic Quadrant for Advanced Analytics Platforms", Gartner report, 09 February 2016.

[5] Sallam, R., Hostmann, B., Schlegel, K., Tapadinhas, J., Parenteau, J., Oestreich, T. "Magic Quadrant for Business Intelligence and Analytics Platforms, Gartner report, 23 February 2015.

[6] T.H. Davenport and J. G. Harris "Competing on Analytics: The New Science of Winning", Harvard Business Press, 2007, 218 p.

[7] Parenteau, J., Sallam, R., Howson, C., Tapadinhas, J., Schlegel, K., Oestreich, T. "Magic Quadrant for Business Intelligence and Analytics Platforms, Gartner report, 04 February 2016.

[8] Hurwitz, J., Kaufman, M. and Bowles, A. Cognitive Computing and Big Data Analytics, John Wiley & Sons, 2015.

[9] Quintero et al., Casali, D., Lima, M., Szabo, I., Maciej Olejniczak, M., Mello, T., Santos, N. IBM Software Defined Infrastructure for Big Data Analytics Workloads, IBM Redbooks, 2015, 178 p.

[10] Siegel, E. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die, John Wiley & Sons, 2013, 320 p.

[11] "Big Data Taxonomy", Cloud Security Alliance, 2014, <https://cloudsecurityalliance.org/research/big-data/> (accessed 15 June 2016).

[12] Poli, R., Healy, M. and Kameas, A. Theory and Applications of Ontology: Computer Applications Set: Theory and Applications of Ontology, XVIII, 576 p. Springer.

[13] Gartner newsroom, <http://www.gartner.com/newsroom/id/3198917> (accessed 15 June 2016).

[14] Domenico, F., Kenneth R., and Barbara, R. 2015. Can oil prices forecast exchange rates? An empirical analysis of the relationship between commodity prices and exchange rates. *Journal of International Money and Finance* 54: 116-141

[15] Elliot, G., and A. Timmermann. 2013. *Handbook of Economic Forecasting*. Elsevier Science and Technology Books, Inc.

[16] Finam. <http://www.finam.ru/analysis/quotes/?0=&t=83156983198917> (accessed 15 June 2016).

[17] Bose, R. "Advanced analytics: opportunities and challenges", *Industrial Management & Data Systems*, Vol. 109 Iss 2, 2009, pp. 155 – 172.

[18] Sherman, R. *Business Intelligence Guidebook: From Data Integration to Analytics*, Morgan Kaufmann Publishers, 2015, 551 p.