

# A Multifaceted Approach to Amazon's Financial Performance: Time Series, Difference in Difference, and Regression Discontinuity Analysis of R&D, Marketing, and Mobile Adoption

*Sarikyan Arsine A.*

*PHD candidate, Mathematical Modeling of Economy,  
Armenian State University of Economics (Yerevan, RA)*

<https://orcid.org/0009-0006-5847-3400>

[asarikyan@gmail.com](mailto:asarikyan@gmail.com)

**UDC:** 330.4; **EDN:** MCLMXN; **JEL:** C5, R15;

**DOI:** 10.58587/18292437-2024.2-116

**Keywords:** Econometric models, ARIMA, SARIMA, ARDL, Difference-in-differences method, Regression discontinuity model

## Բազմակողմանի վերլուծություն Amazon-ի ֆինանսական գործունեության մասին՝ ժամանակային շարքերի, տարբերության տարբերությունների և ռեգրեսիոն վերլուծությունների միջոցով

*Մատիլյան Արսինե Ա.*

*Հայցորդ, Տնտեսության մաթեմատիկական մոդելավորման ամբիոն,  
Հայաստանի պետական տնտեսագիտական համալսարան (Երևան, ՀՀ)*

**Ամփոփագիր.** Ժամանակակից բիզնեսի լանդշաֆտում Ամազոն-ը դարձել առաջնակարգ ուժ՝ հաստատելով իր գերակայությունը համաշխարհային առցանց առևտրի և ամպային հաշվարկների ոլորտներում: Այս վերելքը հիմնված է շարունակական նորարական մոտեցումներով, որը դրսևորվում է հետազոտությունների և զարգացման (Հ&Ձ) հանդեպ զգալի ներդրումներով: Այս ուսումնասիրությունը փորձում է պարզել հիմնական փոփոխականների ազդեցությունը, ինչպիսիք են (Հ&Ձ) և մարքեթինգային ծախսերի և պատմական տվյալների ազդեցությունը Ամազոն-ի եկամուտների վրա:

Բացի այդ, Ամազոն-ի բջջային հավելվածը համատարած ընդունվել է ԱՄՆ-ի սպառողների կողմից և օգտատերերի կեսից ավելին ընդգծում է որ դա իրենց կյանքի անբաժանելի մաս է կազմում: Երևույթ, որը հնարավոր եղավ միայն 2007-ին՝ առաջին սմարթֆոնի ստեղծումից հետո: Այս հետազոտությունը նպատակ ունի հայտնաբերելու Ամազոն-ի ֆինանսական ցուցանիշները ձևավորող գործոնների բարդ փոխազդեցությունը՝ նպաստելով ժամանակակից բիզնեսի դինամիկայի ավելի խորը ըմբռնմանը՝ օգտագործելով տարբեր էկոնոմետրիկ մոդելներ՝ ARIMA, SARIMA և ARDL:

Բացահայտումները ցույց են տալիս, որ ARIMA և SARIMA մոդելները բավականին լավ չեն կանխատուում Ամազոն -ի եկամուտները, մինչդեռ ARDL մոդելն ապահովում է եկամուտների ավելի ճշգրիտ կանխատեսումներ՝ ապահովելով R-քառակուսի արժեք հավասար 62%-ի: Ավելին՝ իրականացնելով տարբերության տարբերությունների մեթոդը, Ամազոն-ը որպես ազդեցության ենթարկվող խումբ, իսկ մրցակիցներից մեկին որպես վերահսկիչ խումբ, ուսումնասիրությունը բացահայտեց սմարթֆոնի թողարկման էական ազդեցությունը Ամազոն-ի եկամուտների վրա: Մասնավորապես, վերլուծությունը ենթադրում է 2007-ից մինչև 2016 թվականների 41,000 միլիոն դոլարի կուտակային էֆեկտ, որը հաստատվում է ռեգրեսիոն դադարման մոդելի միջոցով, ինչը նույնպես հաստատեց հայտնաբերված արդյունքները:

**Հանգուցաբառեր՝** Էկոնոմետրիկ մոդելներ, ARIMA, SARIMA, ARDL, տարբերության տարբերությունների մեթոդ, ռեգրեսիայի դադարման մոդել

## Многогранный подход к финансовым показателям Амазон используя временные ряды, разница в различиях и регрессионный анализ разрывов

*Сарикян Арсине А.*

*Соискатель, кафедра Математическое моделирование экономики,  
Армянский государственный экономический университет (Ереван, РА)*

**Аннотация.** В современном мире бизнеса Амазон стал выдающейся силой, утверждая свое доминирование в сферах электронной коммерции и облачных вычислениях. Это господство подкреплено твердой приверженностью к инновациям, что проявляется в значительных инвестициях в научные исследования и разработки (НИОКР), а также в стратегические маркетинговые усилия. Целью данного исследования является выявление связей ключевых переменных, таких как расходы на НИОКР, расходы на маркетинг и исторические

финансовые данные, на прибыль и формирование доходов Амазона. Кроме того, мобильное приложение Амазона используется повсеместно в США для онлайн-покупок и более чем половина потребителей подчеркивает его важную роль в повседневном использовании. Катализатором этого явления стало революционное появление первого айфона в 2007 году, что значительно увеличило траекторию доходов Амазона. Посредством тщательного анализа это исследование направлено на выяснение сложного взаимодействия факторов, влияющих на финансовые показатели Амазона, используя эконометрические модели, такие как ARIMA, SARIMA и ARDL. Результаты показывают, что, хотя модели ARIMA и SARIMA с трудом прогнозируют доход, ARDL обеспечивает более точные прогнозы, достигая значения R-квадрата 62%. Кроме того, используя метод разницы в различиях, с Амазоном в качестве экспериментальной группы и одним из конкурентов в качестве контрольной группы, исследование выявило существенное влияние запуска смартфона на рост дохода. В частности, анализ предполагает совокупный эффект в размере 41,000 миллионов долларов США с 2007 по 2016 год, что дополнительно подтверждается моделью регрессионного разрыва, которая определяет четкую точку отсечения, совпадающую с датой запуска смартфона.

**Ключевые слова:** эконометрические модели, ARIMA, SARIMA, ARDL, метод разницы в различиях, модель разрыва регрессии

### **Introduction**

In the contemporary landscape, Amazon stands out as the preeminent global entity in both e-commerce and cloud computing realms. Its position as a beacon of innovation is underscored by its strategic and substantial investments in research and development (R&D) and marketing initiatives. As a result, the study endeavors to prognosticate Amazon's Net Income (NI) and Revenue, leveraging variables such as Research and Development expenditure, Marketing expenses, and historical financial data pertaining to the company's past revenue or NI. Additionally, it seeks to discern the extent to which these variables exert a significant influence on the company's overall revenue generation.

Furthermore, in today's digital world, Amazon's mobile application has become integral to the online shopping experiences of over 50% of consumers in the United States. This shift, catalyzed by the introduction of the first iPhone in 2007, has heralded a paradigmatic transformation in the e-commerce sector at large, with Amazon witnessing a substantial impact on its total revenue. The paper delves into the analysis of this pivotal development, elucidating its implications within the broader context of Amazon's revenue dynamics.

This paper aims to quantify the aforementioned effects utilizing econometric models such as ARIMA, SARIMA, and ARDL for Net Income. Furthermore, the study endeavors to capture the influence of the iPhone's launch in 2007 on internet usage and the e-commerce industry, historically limited by mobile phones with restricted internet access, through methodologies like Difference in Differences and Regression Discontinuity.

### **Literature Review**

Marquardt and Murdock [4] found that there is a positive relationship between the change in advertising expense and the change in sales in a cross-section of firms. This was interpreted as the

impact of advertising on sales. Their findings concluded that firms could increase their profits by increasing their levels of advertising.

Concerning the predictive part of the research, ARIMA was implemented, as it explicitly caters to a suite of standard structures in time series data, and as such provides a simple yet powerful method for making skillful time series forecasts.

An example is research in the sphere of telecommunication. As it is stated in the paper, with the increasing competition in the telecommunications industry, the operators try their best to increase telecom income via various measures, one of which is to set an amount of income as a goal to make the encouragement. Since an accurate forecast of income plays an important role in income target setting, the paper builds a time series ARIMA model based on the analysis of income data. The results show that the proposed model fits income data well and performs well in forecasting. Moreover, an accurate income forecasting value was estimated which could help top managers in companies grasp income trends. [6, p. 2560-2563]

Most of the time the earnings forecasts are based on analysts' prospecting of firms' growth and profitability; to forecast earnings, stock analysts shape financial models that estimate prospective revenues and costs, and incorporate other factors like economic growth, currencies and other macroeconomic factors that influence firms' growth [5, p. 60-64].

One example is a study which aimed to estimate and forecast the earnings of the firms listed in Amman Stock exchange (ASE), using a time series data of earning per share (EPS) for the period from 1978 till 2016. The results showed that an ARIMA offers an excellent technique for forecasting any variable like EPS and with this model the forecasting accuracy is high. It was concluded that firms' earnings show a slow

increasing trend for upcoming 38 financial years. [1, p. 378-381]

Another widely used model is the Auto Regressive Distributed Lag Models (ARDL) model, which plays a vital role when comes a need to analyze an economic scenario. An interesting study was done to develop a model for estimating US aggregate hotel industry revenue. It is the first study that adopted the ARDL bound approach and analyzed the predictive capacity of macroeconomic variables on the aggregate hotel industry and sub-segment revenue, showing significant utility in accurately estimating the revenue. [3, p. 8-15]

The research also tries to capture the influence of the launch of the iPhone in 2008 on the internet usage and e-commerce industry. In the past, mobile phones had limited Internet access, however, with the new iPhone, people have the opportunity to browse via a hugely improved device, which can display web pages more like they actually looked on a desktop computer [3].

Moreover, the study tries to capture the difference of eBay and Amazon, thus it was essential to prove their relationship as competitors. According to Krishnamurthy [7, p. 36-42], over time, Amazon started to identify eBay as its direct competitor. A deep analysis shows that these two companies are direct competitors. For example, eBay was the sole site to receive more visitors than Amazon during the 2002 holiday season. In addition, it is now well known that Amazon considers eBay to be its biggest competitor. In addition, historical data shows that growth rate of revenues of the two companies move in opposite directions. This is a convincing fact to assume that the two companies are pure and direct competitors.

### Data Description

The data used is a time series data that includes Amazon's and eBay's quarterly revenue and R&D expenses, as well as Amazon's net income and marketing expense for the period of 1998-2018 and contains a total of 85 observations. The main source of the data is ycharts.com, which is a premier cloud-based investment decision-making platform. However, a short period was missing from the data of marketing expense, which was filled by the data imputation method. Some of the missing values were looked up and filled from Amazon's annual reports. While some of the values that were not available in the official reports of the company were imputed using the Moving Average (MA) method of data imputation. In addition, the internet usage variable was included from the [www.internetworldstats.com](http://www.internetworldstats.com) and include only 50 observations.

All the variables were non-stationary and the simple differencing was not enough for making

them stationary. Therefore, most of the data became stationary after differencing their logarithmic values.

### Methodology

#### ARIMA & SARIMA for Net Income

For making forecasts of AMAZON's quarterly net income (Figure 1), ARIMA and SARIMA models were run as well. The data was non-stationary with a sharp trend, especially close to the last periods.

Figure 1: Amazon Net Income Over Time

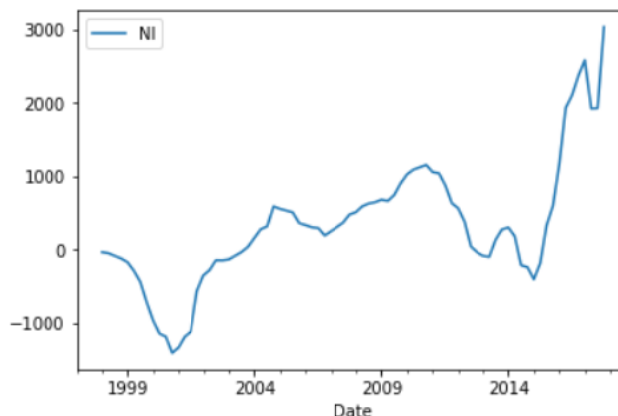
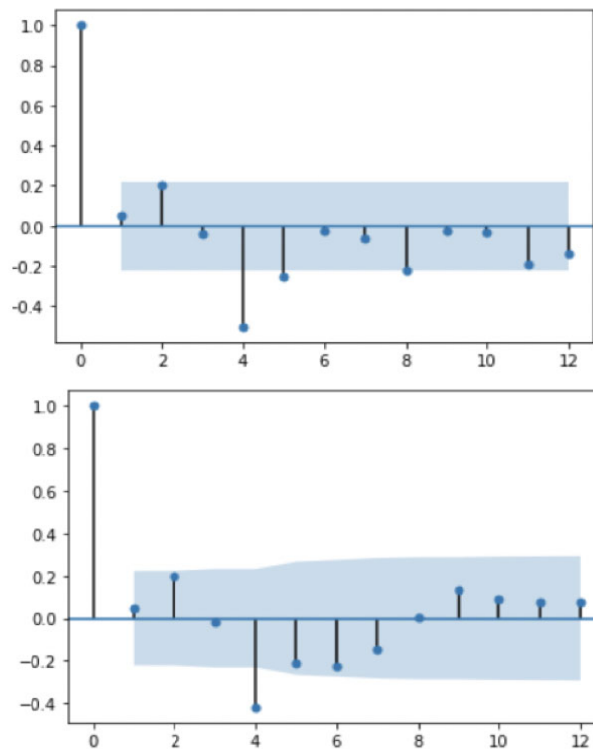


Figure 2: ACF and PACF plots



To make the data stationary the following steps were taken

1. Normalizing the data (the values were rescaled into the range 0 and 1)
2. Making logarithmic differentiation
3. Differentiation of first order.

As a result, the ADF test showed  $p < 0.05$  for Net Income (the data became stationary).

In order to run ARIMA, first, the optimal number of  $p$ ,  $d$ ,  $q$  parameters had to be selected based on the smallest number of AIC and BIC criteria (both were the smallest for the same number of orders). For identifying the optimal number of orders for ARIMA, `auto_arima` was run. The model with order (1,1,0) was chosen. The output showed  $p < 0.05$ , for AR term of Net Income (which underwent the above-mentioned transformations), showing that it is significant at 5 % significance level. However, correlograms (ACF and PACF plots) indicated that there is a serial correlation in the residuals (with the values of 4th lag).

It was supposed that a seasonal pattern may be present in the data. In order to tackle the residuals' autocorrelation problem, SARIMA model with order (1,1,1) and seasonal order (0,0,1,4) was tested.

The results were the following: coefficients for `ar.L1` (AR term at lag 1) and `ma.SL4` (SAR of lag 1) terms were statistically significant with  $p < 0.05$  and `ma.L1` (MA term at lag 1) statistically insignificant. The problem with residuals' autocorrelation was eliminated with all correlation after lag one not significant. Model predictions were compared with the actual ones, as well as MSE was calculated. The result 0.03 can be considered quite good (very close to 0), however, the model will probably not be as accurate for the forecasted data, as it was on the training set.

#### **ARDL (6,0) for Net Income**

ARDL model was run for predicting the net income, as well. All the variables (Net Income, R&D, Marketing expenses) were transformed to comply with the stationarity requirement. The model was run until the residual's autocorrelation problem was resolved. The optimal number of lags for NI was chosen 6. OLS regression was run, with the output of  $R^2$  equal approximately to 62%, and the coefficients of Research and development and Marketing expenses statistically significant. The model accuracy was again checked MSE, exposing a result 0.02.

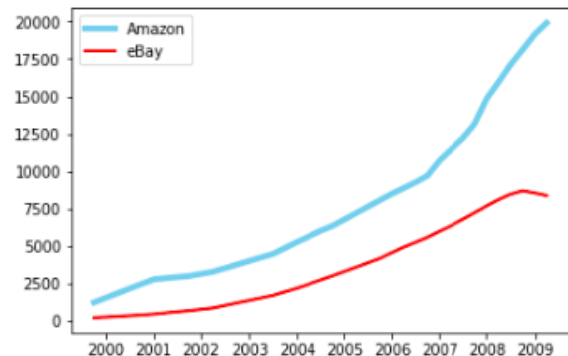
#### **Difference in Differences**

One of the interesting things that the research found during the research about Amazon and its revenue influencers, was the launch of the first iPhone on June 29, 2007. The first iPhone not only triggered the exponential growth of the internet, which enables customers to experience the full benefit of connectivity directly from their pockets, but also skyrocket the growth of Amazon's revenue. Even more interesting is that the study was not able to find similar information regarding eBay, which is the direct competitor of Amazon. As a result, it tried to implement the method of difference in differences, by proving that eBay can be considered

the control group. In order to validate this assumption, several facts need to be proved;

1. There should be a parallel relationship between Amazon's and eBay's Revenue
  2. The internet usage variable should not be significant for determining eBay's revenue
  3. The internet usage variable should be significant for determining Amazon's revenue
- 1) The parallel relationship can be checked by simple plotting of two revenue lines (Figure 3);

**Figure 3:** Amazon and eBay Revenue Over Time



It is easy to notice that two lines are similar and more or less are parallel until the mid of 2007, exactly when the iPhone was launched, thus the first assumption is validated

2) Then it is needed to construct a model for checking the relationship. As all of the interested variables were non-stationary, the modifications, such as differencing and logarithms differencing was used for making them stationary. Afterwards, the ARDL (4,0) model was created, with the R-square equal to 86% and internet usage variable being non-significant, as it has a p-value (0.885) higher than 0.05. Therefore, it is reasonable to conclude that internet usage did not affect eBay's revenue, so it can be considered as a control group.

3) The cointegration test was conducted for Amazon's logarithm variables, yet neither the revenue and advertising expense, nor the revenue and research and development are cointegrated, as their p-values are higher than the 0.5. The ARDL (2,0) model resulted in the R-square equal to 52% and internet usage is having p-value equal to 0.524. After changing the model to the ARDL (1,0) and removing research and development parameter, the model resulted in the R-square equal to 31% and internet variable is significant. Even though the second model has lower R-square it still gives the possibility to conclude that under certain conditions internet usage can have a significant impact on Amazon's revenue.

The first difference was about the change of Amazon's revenue before and after the launch date and the second difference was eBay's revenue change as a control variable. All the differences were tested with t-test and proved to be significant. The total effect on Amazon's revenue can be estimated by subtracting eBay's change from Amazon's one, which is equal to \$41,000 million, till March 2016.

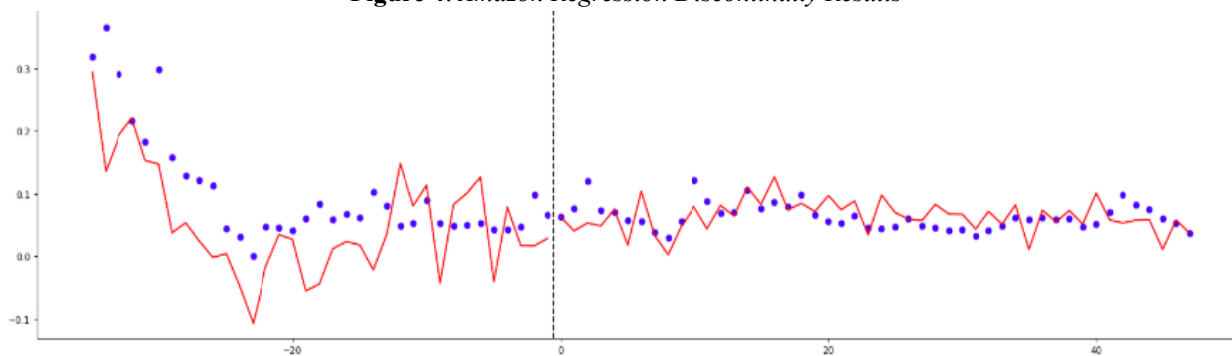
### Regression Discontinuity

After the results of the difference in differences, it was decided to explore the launch

data as a cutoff point for the revenue by making data stationary and creating a simple OLS model. Two models were tested, one with the marketing variable and one without. The models have almost the same R square (66.3% and 64.8%), yet the AIC and BIC scores were significantly smaller for the model without marketing. Therefore, it was decided to proceed with that model.

Please note that the new dummy variable was added; 0- before the iPhone launch and 1-after the launch. Regression discontinuity graph clearly showed a cutoff point;

**Figure 4: Amazon Regression Discontinuity Results**



### ARDL (3,0) for Revenue

The new ARDL (3,0) model was created for making predictions. Several ARDL models were tested and the ARDL (3,0) model was selected, as it has the highest R-square equal to 62%.

The results show that the lag\_4 is in the border and there is an autocorrelation at lag\_12. After getting the predicted results for the period, which was not included in the model evaluation, as the model overestimates the predicted values;

### Results

The first part of the report was about identifying the effect of the first iPhone launch on the revenue of Amazon through the difference in differences model and regression discontinuity. Two models demonstrate that the launch has a significant impact, more specifically it resulted in an increase of \$41,000 million for Amazon, till March 2016.

In addition, the ARDL (3,0) model was developed as the best one for predicted Amazon's Revenue. The model has a score of 62% and the predicted values need to undergo some transformations for becoming representative of actual revenues. It is possible that the transformed values would be biased and not accurate as of the base for adding logarithmic differences is chosen from the actual model, which makes it strongly dependent on historic events. The second part was about identifying the best model for making predictions. The ARDL (2,0), ARIMA, and

SARIMA were tested, which were not good as the residual autocorrelation was still present.

### Limitations and Constraints

The most apparent limitation of the analysis is the small number of variables and observations. The analysis emphasizes the effect on the revenue, which in reality depends on the number of things, other than marketing and research and development.

Interestingly, 2007-2009 was a very productive and important year for the Amazon, as besides the iPhone launch

- Amazon introduced the Kindle (2007), which at that time had a starting price of \$399
- Acquired audiobook company Audible (2008) and shoe shopping site Zappos (2009)

Moreover, the number of observations was pretty small and can possible overfit the data. The limitation is even more strict in the analysis of the difference in differences method, as there the missing values of internet usage were dropped, which resulted in only the 50 observations.

### Conclusion

In summary, our analysis indicates that both ARIMA and SARIMA models are inadequate for accurately predicting Amazon's Net Income due to unresolved autocorrelation issues within the residuals. Conversely, the ARDL model demonstrates superior performance in forecasting the company's revenue, exhibiting no

autocorrelation and achieving an R-square value of 62%.

Furthermore, employing the difference-in-differences method with Amazon as the treatment group and eBay as the control, our study reveals a significant impact of the iPhone launch on Amazon's total revenue. Specifically, the analysis suggests a cumulative effect of \$41,000 million over the period spanning 2007 to 2016. Additionally, the launch effect was corroborated through the regression discontinuity model, which identified a distinct cutoff point coinciding with the launch date.

#### Reference

1. **Alsinglawi, Omar & Aladwan, Mohammad & Bouqalieh, Bassam & Alwadi, Sadam.** (2016). Forecasting Earnings of Firm's Listed in ASE Using ARIMA Model, [https://www.academia.edu/38386103/Forecasting\\_Earnings\\_of\\_Firms\\_Listed\\_in\\_ASE\\_Using\\_ARIMA\\_Model](https://www.academia.edu/38386103/Forecasting_Earnings_of_Firms_Listed_in_ASE_Using_ARIMA_Model) creation. Financial Analysts Journal, 56-67
2. **Han Chen, Rui Chen, Shaniel Bernard, Imran Rahman,** (2019) "US hotel industry revenue: an ARDL bounds testing approach", International Journal of Contemporary Hospitality Management,
3. **Kelly, H.** (2012, June 30). 5 ways the iPhone changed our lives. Retrieved from <https://edition.cnn.com/2012/06/28/tech/mobile/iphon-e-5-years-anniversary/index.html>
4. **Marquardt, Raymond A. and Gene W. Murdock** 1984. "The Sales/Advertising Relationship: An Investigation of Correlations and Consistency in Supermarkets and Department Stores," Journal of Advertising Research
5. **Ramezani, C. A., Luc Soenen, & Jung, A.** (2002). Growth, Corporate Profitability, and Value Creation. Financial Analysts Journal, 58(6), 56–67. <http://www.jstor.org/stable/4480435>
6. **Wang M, Wang Y, Wang X, Wei Z** (2015) Forecast and analyze the Telecom Income based on ARIMA model. Open Cybern Syst J 9(1):2559–2564, <https://pdfs.semanticscholar.org/d1d6/4294a1a557f1bcf800ad5d748b1a1bdaa7ba.pdf> +
7. **Krishnamurthy, S.** (n.d.). A Comparative Analysis of eBay and Amazon. Intelligent Enterprises of the 21st Century, 29-44. doi:10.4018/978-1-59140-160-5.ch003

*Сдана/Հանձնվել է՝ 27.02.2024*

*Рецензирована/Գրախոսվել է՝ 05.03.2024*

*Принята/Ընդունվել է՝ 12.03.2024*