Forecasting FDI inflow into the US using Vector Autoregression

Guy Jansen

Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China
Email: 612371gj@eur.nl

Abstract

This paper investigates Foreign Direct Investment (FDI) inflow in the United States, employing Autoregressive (AR) and Vector Autoregressive (VAR) models to predict future inflows. Analysing historical patterns and key drivers, the AR model demonstrates commendable short-term predictive power, while the VAR model, incorporating dynamic relationships between factors, significantly improves accuracy in the long term. The study reveals a likely continuation of historical FDI growth trends, with short-term fluctuations attributed to wage index escalation and trade stagnation. Post-2028, a promising trajectory for FDI inflows is anticipated, supported by the stabilizing influence of the wage index. Overall, this research enhances understanding of FDI dynamics, offering valuable insights for investors and researchers interested in the future of FDI in the United States.

Keywords

FDI; United States; Vector Autoregression; VAR; AR

1. Introduction

In an era characterised by globalisation and international connectivity, Foreign Direct Investment (FDI) is one of the most important drivers of the global economic landscape. The United States, being the largest economy in today's world (per GDP), attracts a substantial FDI while concurrently investing in other economies. understanding the intricate patterns, key drivers and forecasting FDI trends in the future is an imperative undertaking with profound implications for investors and researchers.

1.1. Background

Foreign Direct Investment defines a cross-border investment made by multinational enterprises, individuals or governments by acquiring foreign assets with the prospect of lasting interest and/or influence on foreign enterprises. These investments can take on many forms such as mergers, capital infusions and the establishment of new ventures. As world's biggest economy, the United States also remains the country with the highest FDI inflow. The objectives for these investments are multifaceted and can range from geopolitical influence to future economic gain.

1.2. Research objectives

The main objective of this research endeavour is to dive deep into the nuanced patterns, drivers and factors that play a key role in FDI and predict future trends in FDI inflow based on these findings. Therefore, the objectives can be broken down to:

1. Analysing Drivers, finding the factors that influence FDI activities, ranging from macroeconomic indicators to regulatory frameworks and other considerations.
2. Developing Projections, using a regression model to forecast future growth to provide valuable insights into the future of FDI in the United States.
3. Analysing the results, by comparing the different models with each other, as well as an OLS model the overall fit of the model can be analysed.

1.3. Research questions

To guide this exploration, the following questions are posed as the foundation of this paper,

1. What are significant historical patterns on FDI inflow to the United States?
2. What key drivers can explain the patterns found in the historical data?
3. How can we accurately forecast the future FDI inflow to the United States?

With these questions addressed, this study can offer a comprehensive analysis to further strengthen the current understanding of FDI inflow of the United States, providing insights into global economic dynamics and the development.

2. Literature review

The aggregate amount of FDI inflow to the United States, and the origin of this capital, is influenced by a myriad of factors. There are numerous research papers that have already determined important factors that influence FDI inflow with respect to countries or regions, these will now be discussed in a brief overview of the literature relevant to this study.

In many research papers the most important factor in determining FDI inflows is the host countries market size (Moosa, 2002). This theory is also backed by the research paper of Dr. R. Grosse, et al. (1996), who analysed the FDI outflow of the United States by country of origin. By investing in larger markets, foreign investors will be exposed to the benefits of a large-scale economy, with the result of increased sales. Another positive factor that plays a key role was the host country's export to the investing country. Besides these positive factors, there are factors with a negative correlation to FDI inflow as well. Grosse's study found a significant negative causation between the host country's imports from the investing country, the cultural and geographic distances of the host country to the investing country and the exchange rate between the investing country and the host country. As around 60% of FDI inflow to the US is contributed by European countries these findings seem to match the current situation well, as the United States is very favourable place for foreign entities to invest in.

Another important relationship identified in previous studies is between trade and FDI (Billington, 1999, Wijeweera, et al., 2008). Billington concludes that the level of trade (import + export of a country) between the host country and the rest of the world is correlated with the amount of FDI inflow to the host country. Using the Granger-causality method, a linear relationship between trade and FDI was uncovered that shows that 50% of the observed linear feedback from FDI to Trade Openness can be explained by previous changes in FDI (Aizenman, 2006). Similarly, the same study found that 31% of the linear feedback from Trade to FDI can be explained by the Granger-causality.

The difference in wage rate is another important factor determined in FDI inflow by previous research. In a study conducted by Cushman (1987) the results showed a significant relation between the wage gap between investing countries and the host country. Thus, an increase in the wages in an investing country, or decrease in wage rates for the host country result in a higher FDI inflow.

The last important factor identified in previous studies that showed to have a significant effect on FDI was the exchange rate. Cushman (1988) found that the expected appreciation of the dollar is highly correlated with reductions in FDI inflows into the United States. The implications of actual depreciation of the exchange rate where researched by L. Goldberg (2006) and found that because the exchange rate affects the relative wealth of the investing country has opposed to the host country, the wage and production costs are for the foreign country are reduced relatively. Thus, the host country has become more attractive to FDI.

There is a general consensus that interest rates are negatively correlated with FDI inflow due to the impact it has on the cost of capital and therefore also on the return on investment. This theory is supported by many empirical studies (Sahiti, et al. 2020, Siddiqui, et al., 2014). Although this may be the general consensus, another study involving 84 countries found no significant impact of interest rates on the FDI inflows because investments where direct, and not portfolio investments (Li, Liu, 2005). However, the research conducted by Scott-Green (1999) has compelling findings on the negative correlation between the host countries interest rate, relative to foreign investing countries. The author does note limitations to the analysis, as there is an absence of disinvestment data resulting in a potentially biased outcome. Despite this limitation, the expected relationship with the relative interest rate variable is noted to be in line with financial theory.

3. Methodology

3.1. Data collection

In pursuit of a comprehensive understanding of Foreign Direct Investment inflows and to ensure accurate findings, rigorous data collection methods were used to gather all relevant information. To make sure the data was reliable and consistent, the most used data source was the United States federal government. Specifically, the United States Bureau of Economic Analysis (BEA), and the Federal Reserve Economic Data (FRED) were mainly used to conduct this research. Data quality checks and validation processes were implemented to make sure the dataset was valid and robust.

3.2. Sources of FDI inflow data

The actual data on the FDI inflow into the United States came from BEA. BEA provided a dataset with the FDI inflow per country and region. As a result, it is clear to see which countries and regions contributed to the FDI inflow the most. One of the main targets of this research is to uncover patterns within this data to anticipate the future FDI inflow, and where it is coming from.
Besides relevant data on the FDI inflow, macroeconomic data is used to look at the correlation and effects it has on the FDI inflow. All macroeconomic data comes from the Federal Reserve Economic Data, with the exception being the National Average Wage Index. The data on this index was not available with FRED however, it is listed with the United States Social Security Administration (SSA).

3.3. Timeframe of the study

The timeframe to which this study is conducted extends from 2000 to 2022, encompassing the last two decades of FDI history in the United States. The choice for the particular timeframe was motivated by the significant global economic shifts and events that took place inside this timeframe. The last two decades have been characterised by rapid technological innovations, globalisation and tremendous changes in international trade dynamics. The major events that unfolded inside this timeframe e.g., the collapse of the dot-com bubble, the 2008 global financial crisis and the Covid-19 pandemic have all shaped the landscape of Foreign Direct Investments globally. Within the chosen timeframe, this research aims to capture the intricate connections between economic, and technological factors that have influenced investments patterns. Additionally, this timeframe is best to complete the main objective behind this study, predicting the future FDI inflow into the United States. As this is the most recent data available it is most comparable to the current macroeconomic situation, so it holds the most valuable data for predicting future FDI inflow. In conclusion, by using the timeframe 2000 to 2022, it ensures a comprehensive examination of the evolution of FDI inflow, allowing for a nuanced understanding of the trends and drivers while still being recent enough to derive predictions about the future of FDI inflow in the United States.

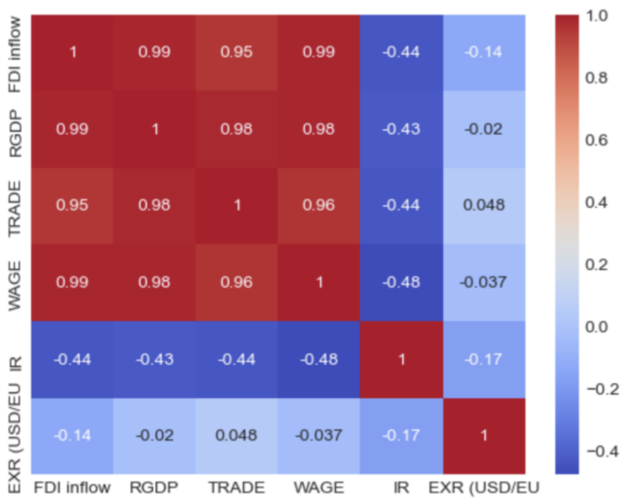
3.4. Data analysis techniques

To draw meaningful conclusions, FDI inflows need to be assessed with a robust analytical framework. Quantitative data analysis techniques will be used to provide a comprehensive understanding of the macroeconomic factors that influence FDI inflow. With a time series analysis, the data is analysed to look for any temporal trends, potential patterns, cycles and anomalies. This time series analysis will consist of a VAR analysis to uncover the relationships between FDI inflow and macroeconomic indicators, leading to the identification of the most significant determinants of FDI inflow. These determinants will then be used to forecast future FDI inflow. The fit of this model will be benchmarked against an Ordinary Least Squares (OLS) model in order to assess if there is a significant improvement in the accuracy of the model by incorporating the dynamic relationship between the factors.

4. Descriptive statistics for FDI factors & Correlation, 2000-2022

	FDI inflow	RGDP	TRADE	WAGE	IR	EXR (USD/EU
count	2.300000e+01	23.000000	23.000000	23.000000	23.000000	23.000000
mean	2.771898e+06	17751.800174	4451.444174	44336.990870	1.447391	1.213935
std	1.275165e+06	2354.976108	931.208512	8830.422812	1.855386	0.151694
min	1.256867e+06	14229.765000	2818.730000	32154.820000	0.070000	0.890100
25%	1.737292e+06	16311.042000	3836.286000	37802.175000	0.120000	1.127250
50%	2.433848e+06	17222.583000	4523.394000	42979.610000	0.540000	1.210100
75%	3.757061e+06	19593.352000	5209.066500	49482.020000	1.990000	1.323300
max	5.254816e+06	21847.602000	5849.272000	63795.130000	6.400000	1.460300

Correlation Matrix between FDI factors



5. Autoregressive Model (AR) model

The first method of forecasting the future inflow of FDI into the United States that will be used is running a simple Autoregressive Model. in this type of regression model values are modelled as a linear combination of its based values. Mathematically this formula is denoted as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t$$

With y_t being the current value, ϕ the autoregressive coefficient and ϵ_t the error term. As for the data source, the total sum of all FDI inflows into the United States was used. As the time series data used in this research ranges from 2000 up until 2022 there are 23 different values in this dataset. For the AR Model this dataset has been split into a 75% training and 25% testing set. Hence, the forecasted FDI inflow starts in 2018. To assess the accuracy of this model the Mean Squared Error and the R^2 Score are used.

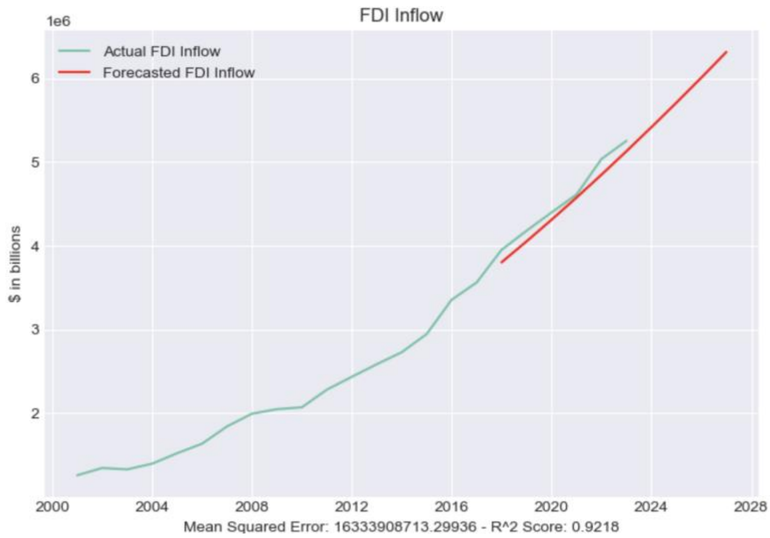


Fig 5.1 - result of the Autoregressive Model on the Original Data

In Fig 2.1 the results of the AR can be assessed. The autoregression predicts FDI inflow will increase with a yearly coefficient of approximately 1.064. The plot shows that AR has forecasted FDI inflow for the test data (2018 to 2022) quite accurately. The Mean Squared Error however is still relatively high which would mean the model did not give a very accurate prediction. To further assess the results the R^2 Score was also used. With a score of 0.9218 this would be considered a very good score, especially in the context of an AR. The R^2 Score means that approximately 92.18% of the variability can be explained by the independent variable in the AR. And thus, it captures a large portion of the underlying patterns inside the data. As the regression can only produce a linear forecast this prediction should only be used to forecast FDI on a short timeframe as there appears to be exponential growth.

6. Vector Autoregression (VAR) model

The amount of FDI inflow into the United States is indubitably dependent on a myriad of factors. Though, the factors affect each other making it hard to accurately forecast FDI inflow only using an autoregressive model, as

they are only designed to capture the autocorrelation within a single variable and do not take the influence between variables into account. Because of this dynamic relationship between factors, a Vector Autoregressive model is a good alternative that can be used in order to take these relationships into account. Another reason for the use of a VAR model is the absence of need for exogenous variables, as all variables are considered endogenous. This helps simplify the modelling process and avoids the need to explicitly define the relationship between endogenous en exogenous variables. As discussed in the literature review, the factors that have been selected for this model have shown to have a significant effect on FDI inflow. In order to get an accurate forecast the data needs to be stationary.
Stationarity in time series data means that statistical properties, such as the mean and variance remain constant over time. The reason behind this is that it is a more robust way to forecast certain trends as the behaviour is more consistent over time (R. Rosca, 2010).

6.1. Ordinary Least Squares

OLS Regression Results

Dep. Variable:	FDI inflow	R-squared:	0.994
Model:	OLS	Adj. R-squared:	0.992
Method:	Least Squares	F-statistic:	581.6
Date:	Wed, 29 Nov 2023	Prob (F-statistic):	2.25e-18
Time:	08:18:03	Log-Likelihood:	-296.27
No. Observations:	23	AIC:	604.5
Df Residuals:	17	BIC:	611.4
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
RGDP	369.6674	89.366	4.137	0.001	181.121	558.214
TRADE	-205.7359	133.975	-1.536	0.143	-488.398	76.926
WAGE	64.5065	17.133	3.765	0.002	28.358	100.655
IR	-1.322e+04	1.56e+04	-0.847	0.409	-4.62e+04	1.97e+04
EXR (USD/EU	-8.509e+05	1.68e+05	-5.068	0.000	-1.21e+06	-4.97e+05
const	-4.683e+06	6.06e+05	-7.729	0.000	-5.96e+06	-3.4e+06

Omnibus:	0.465	Durbin-Watson:	1.739
Prob(Omnibus):	0.792	Jarque-Bera (JB):	0.178
Skew:	0.212	Prob(JB):	0.915
Kurtosis:	2.917	Cond. No.	1.29e+06

Figure 6.1 – OLS Regression results

To measure how good the model 'fits' the data it will be benchmarked against an Ordinary Least Square (OLS) regression model. The same variables will be used in this model but the relationships between them will not be considered. Figure 6.1 shows the results of the OLS model. With an adjusted R^2 of 0.992 the model seems to perform better than the AR model that did not incorporate any factors, but it must be noted that a model with a very high R^2 score may not generalise well to new data and thus is prone to overfitting. The VAR model will therefore be evaluated based on the Akaike Information Criterion (AIC). The most important reason behind this choice being that the R^2 score does not take the complexity of a model into account. The AIC does take this into account and penalises models with a large number of parameters to make sure the model is not overfitted. Mathematically the AIC is calculated as: $\mathrm{AIC} = 2k - 2\ln(\hat{L})$. With k being the number of parameters (complexity of the model), and L the log-likelihood. The log-likelihood is a measure of how well a model explains the observed data. The AIC should thus be as low as possible, as the best model would be the least complex with the most amount of log-likelihood. As stated in figure 6.1, the AIC of this model is 604.5.

To accurately forecast FDI inflow, the first thing that needs to be done is check whether the time series data is stationary. This is calculated with the Augmented Dickey-Fuller (ADF) test. ADF tests the null hypothesis that a unit root (a feature of stochastic processes that can cause problems in statistical processes involving time series models) is present in a time series sample. The alternative hypotheses in this case being trend-stationarity. Mathematically the ADF test is depicted as: $\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1}$

α being a constant
β the coefficient on a time trend
p the lag order of the autoregressive process
γy_{t-1} the autoregressive component
$\delta \Delta y_{t-1} + \ldots + \delta \Delta y_{t-1}$ the moving average component

6.2. Results Augmented Dickey-Fuller test

Hypothesis, H_0: The time series has a unit root, indicating non-stationarity.

ADF Test for Original Data:

- ADF Statistic: 3.9210381079561674
- p-value: 1.0
- Critical Values: {'1%': -3.769732625845229, '5%': -3.005425537190083, '10%': -2.6425009917355373}

Conclusion: Fail to reject the null hypothesis (Data is non-stationary)

Fig 6.2 - result of the ADF test on the Original Data (FDI inflow)

Let's breakdown the results, the ADF statistic is a measure that looks at how much the data needs to be differenced in order to achieve stationarity. This value should be as low as possible in order reject the null hypothesis. The p-value measures for statistical significance, a value lower then 0.05 will be regarded as statistically significant and the null hypothesis will be rejected.
The Critical Values refer to the thresholds used to interpret the ADF statistic. These are the bounds to see if there is a statistical significance.
By looking at the results of the Original Data there is a notable trend in the data which indicates non-stationarity. In the current state the data cannot be used in a Vector Autoregressive model and thus needs to be differenced. In the context of a time series analysis, differenced data is calculated by subtracting each value from the value preceding it, with as a result a new dataset that contains the changes between consecutive observations. Mathematically, this is defined as: $y_t' = y_t - y_{t-1}$. Differencing is applied iteratively until every variable within the time series is stationary. The final ADF test results show that all data is now stationary.

6.3. Optimal Lag Order

Now that every variable is stationary the optimal Lag Order needs to be calculated. The Lag Order refers to the number of past time points that are used to forecast a current time point. Choosing the optimal lag order is crucial for building an accurate and parsimonious model. Using a Lag Order that is either too high or too low will result in over- or underfitting. To calculate the optimal amount of Lag Order the AIC is used once again. As the results in figure 6.3 of the regression show, the optimal Lag Order for this model is 1. Note that the AIC score with Lag Order 1 is 52.3346. This shows that the VAR model is significantly better at explaining the observed data taking the complexity of the model into account.

```
Optimal Lag Order: 1
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
Date:           Wed, 29, Nov, 2023
Time:                     18:13:44
--------------------------------------------------------------------
No. of Equations:         5.00000    BIC:                    53.8223
Nobs:                     22.0000    HQIC:                   52.6850
Log likelihood:          -701.763    FPE:                5.74592e+22
AIC:                      52.3346    Det(Omega_mle):     1.72061e+22
--------------------------------------------------------------------
```

Figure 6.3 - Optimal Lag Order Results

6.4. Forecasting results

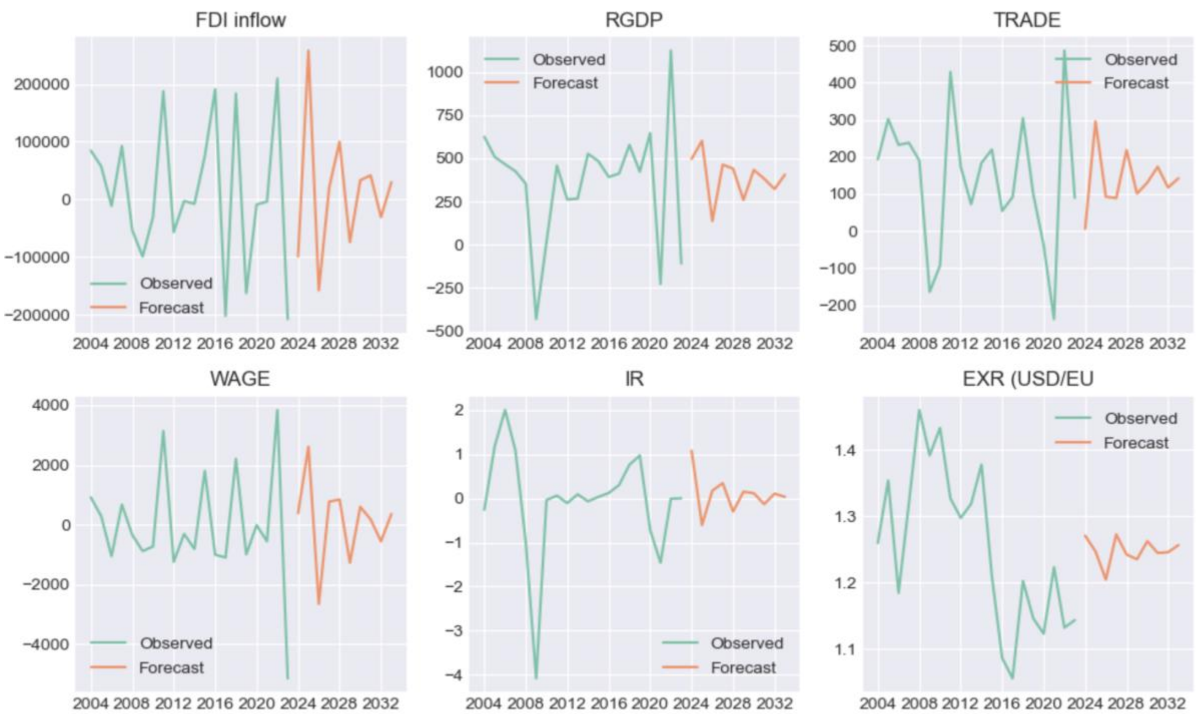


Figure 6.4 - Differenced VAR forecasting results

With the optimal lag order and stationary data it is now possible to forecast the future FDI inflow using the VAR model. In Figure 6.4 the results of the forecast of the stationary data are shown. As this the variables in this dataset are differenced to different degrees it is not possible to draw any conclusions from this yet.

Figure 6.5 - VAR forecasting results normal data

The differenced variables can be reverted back to the original values using the cumsum() function. This is the exact opposite of the differencing function as it takes the cumulative sum of elements along a given axis. Figure 6.5 shows the actual forecasting of the variables over the course of 10 years. The results of this VAR model look quite acceptable as FDI inflow, trade, RGDP and Wage are expected to trend upwards based on the historic data. The Interest Rate however does not seem to be accurate as it becomes negative. As the main goal for this study was to accurately predict FDI inflow while keeping the number of parameters as low as possible to prevent overfitting, there are limitations to this model and its results. The factors that where selected for this study are based on their effects on the FDI Inflow. The interest rate, and the other selected factors are dependent on numerous other complex factors not discussed within this model, so the predictions from this model on the factors outside of FDI inflow should therefore be interpreted with caution, although the general trends they follow might be forecasted correctly (outside of the Interest Rate). Another note to be made is the stagnation of the wage index. Historical data shows that the wage index has never before stayed this stagnated for a longer period of time. This once again shows the constraints inherent in this VAR model, concurrently affirming the potential truth of the overall trends depicted by these factors. Consequently, the model exhibits a measure of explanatory efficacy regarding the prospective trajectory of Foreign Direct Investment (FDI) inflow into the United States but should be used cautiously.

7. Conclusion

The conducted study employed two distinct forecasting models, namely the Autoregressive Model (AR) and the Vector Autoregressive Model (VAR), to forecast Foreign Direct Investment (FDI) inflows into the United States. The AR model, reliant exclusively on historical data of the aggregated FDI inflow of the United States, yielded a reasonably accurate linear prediction. Despite the historical indication of FDI inflows exhibiting a somewhat exponential growth, the AR model demonstrated commendable explanatory power for the short term, as evidenced by an R2 score of 0.9218. In accordance with previous research, a way to enhance the model's explanatory prowess is to incorporate a multitude of factors and take their dynamic interrelations into account. Consequently, the VAR model was introduced. Following confirmation of data stationarity through the Augmented Dickey-Fuller (ADF) test, the optimal Lag Order was obtained via evaluation of the Akaike Information Criterion (AIC). With these parameters optimised, the VAR model forecasted the variables. The results of the VAR model were benchmarked against an Ordinary Least Square Model, which exclusively considers the linear relationship between the target variable and other factors. Against this benchmark, the VAR model substantially improved the AIC from 604.5 (OLS) to 52.33 (VAR), showing that by including the dynamic relationships of the variables into the model, the performance increased significantly.

The primary conclusion that can be drawn from the projected FDI inflow is a likely continuation of historical growth trends. Nonetheless, the model forecasts a diminished growth rate in the short term while anticipating a promising trajectory for FDI inflows into the United States in the long run. A plausible explanation for this short-term projection lies in the escalating trajectory of the wage index, combined with the stagnation of trade and Real Gross Domestic Product (RGDP). As substantiated by the dataset and corroborated by the research of Billington (1999), trade demonstrates a positive correlation with FDI inflows. Hence, trade stagnation, coupled with the adverse impacts of rising wages, is anticipated to influence FDI inflows. A prospective explanation for the anticipated growth surge post-2028 is the stabilisation of the wage index. After 2028, the wage index remains fairly constant around 70.000 while other positively correlated factors increase. Based on the literature reviewed, this would imply an increase in the overall FDI which is the case according to the model.

References

Moosa, I. (2002, januari). *Foreign Direct Investment Theory, Evidence and Practice*. Researchgate. https://www.researchgate.net/publication/265453721_Foreign_Direct_Investment_Theory_Evidence_and_Practice

Grosse, R., & Trevino, L. J. (1996, March 1). *Foreign direct investment in the United States: An analysis by country of origin - journal of international business studies*. SpringerLink. https://link.springer.com/article/10.1057/palgrave.jibs.8490129

WIJEWEERA, Albert* MOUNTER, Stuart . (2008). *A VAR analysis on the determinants of FDI inflows: The case of sri lanka*. USC Gal. https://www.usc.gal/economet/reviews/aeid8116.pdf

NICHOLAS BILLINGTON. (2019, July 2). *The location of Foreign Direct Investment: An empirical analysis*. Taylor and Francis. https://www.tandfonline.com/doi/abs/10.1080/00036846.2019.12067087

Joshua Aizenman, Ilan Noy . (2006, May 19). *FDI and trade-Two-way linkages?*. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S1062976906000305

David O. Cushman. (1987, July 1). *The effects of real wages and labor productivity on foreign ... - JSTOR*. JSTOR. https://www.jstor.org/stable/1058813

Cushman, D. O. (1988, June). *Exchange-rate uncertainty and foreign direct investment in the United States - Review of World Economics*. SpringerLink. https://link.springer.com/article/10.1007/BF02706782

Linda S. Goldberg. (2006). *Exchange rates and foreign direct investment - Federal Reserve Bank of ...* newyorkfed. https://www.newyorkfed.org/medialibrary/media/research/economists/goldberg/ERandFDIArticleGoldberg.pdf

Sahiti, A., Sahiti Ramushi, A., & Zeqiri, R. (2020, August 6). *The determinants of foreign direct investment: Empirical evidence from Kosovo*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3647565

Hira Aijaz, A. Siddiqui. (2014). *[PDF] role of interest rate in attracting the FDI: Study on ASEAN 5 economy*. https://www.semanticscholar.org/paper/ROLE-OF-INTEREST-RATE-IN-ATTRACTING-THE-FDI%3A-STUDY-Aijaz-Siddiqui/7bafc846183672d1f2402203561adb5a8800b1b1.

Xiaoying Li, Xiaming Liu. (2005, January 26). *Foreign Direct Investment and economic growth: An increasingly endogenous relationship*. World Development. https://www.sciencedirect.com/science/article/pii/S0305750X04002013

Susan Scott-Green, Jeremy Clegg. (2002, December 16). *The determinants of new FDI capital flows into ... - wiley online library*. Wiley. https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-5965.00198

Elisabeta R. ROŞCA. (2010, November). *STATIONARY AND NON-STATIONARY TIME SERIES* . The Annals of The "Ştefan cel Mare" University of Suceava. Fascicle of The Faculty of Economics and Public Administration. http://www.annals.seap.usv.ro/index.php/annals/article/view/262/270