"Comparative Analysis of GDP Growth Determinants in the Midwest and Northeast United States: An Empirical Study"

by Aman Dongre

Abstract :

In this empirical study, I would like to discuss, compare and investigate about what factors(variables) would affect the GDP growth rate most for the two regions of the US which are the Midwest and Northeast United States. The Midwest peeked my interest because it's the region where I reside and Northeast as its one of the most popular region in the US which includes the New York state.

We would discuss how these variable will cause the growth rate to increase more in the subsequent years than the others and what caused reasons can cause this to happen. We'll see changes in these variables according to the economic changes which would also cause the growth rate of the economy to change. Also, we will notice if some other factors could affect the GDP growth rate and whether they should be included in the model. This paper will provide valuable insights into understanding the dynamics of economic growth at the national level.

Introduction :

Economic research has long been interested in the relationship between different macroeconomic parameters and economic growth. Through an empirical analysis, we examine in this paper how various factors such as Rent, Financial Institutions, Healthcare, Real estate, Retail trade, and Social assistance affect economic growth. We seek to determine the strength and direction of these associations using econometric approaches to further our understanding of the factors that influence economic growth. The results of this study have significance for practitioners and policymakers to make targeted economic policies, Resource Allocation, Regional development programs, Investment strategies with creative ways to improve living standards and enhance economic stability to support sustainable economic development in those regions.

Literature Review :

https://jmparman.people.wm.edu/research-files/long-run-analysis-of-regional-inequalities-OREP-2021.pdf - discusses the **Long-run analysis of regional inequalities in the US** by Trevon Logan, Bradley Hardy, and John Parman It discusses about the development of a national and fully integrated economy have accelerated and decelerated over time, and these features are driven by unique sets of historical circumstances and policy choices.

Similarities :

- Both studies discuss about the factors that affect the GDP and compare the development of different regions in the US although the factors that affect the GDP are different.
- Both the studies discuss about how historical and contemporary policies have contributed to regional economic disparities, providing insights for policymakers to address these challenges.

Although the deciding factor, The independent variables in our study are the sectors which have contributed to the GDP growth rate and in this study the independent variables are the factors that contribute more to the GDP.

Econometric Models and Methodology Used :

In our study for the GDP growth rate for the 2 regions, the econometric models in EViews would be **ARMA** models which are used to analyze time series data. They also help to capture the autocorrelation of the data and OLS models with Operational Least square methodology along with **forecasting**, **Serial correlations Testing**, **Wald test** for Hypothesis Testing, **Heteroskedasticity** Tests, Stationary or non – stationary variables and ARMA maximum Likelihood tests.

Empirical Analysis :

In our econometric model for GDP growth rate, we will use the sample size from 2018q1 to 2023q4 where q represents the quarters in a year from 1 to 4 which is a **time-series** data model. The data is obtained from the FRED economic data | St. Louis FED. (https://fred.stlouisfed.org/) , For GDP growth rate - https://fred.stlouisfed.org/series/USAGDPRQPSMEI

We would be doing a side by side analysis on the two regions – The Midwest and the Northeast region of the United States with the states for the Midwest being (Missouri, Illinois, Ohio and Wisconsin) and for the Northeast (New York, Pennsylvania, New Jersey and Massachusetts) and look for the sectors that which sector from the past few years is affecting these regions more than the other overall in the region.

The dependent variable for this model would be the GDP growth rate and the independent variables would be the **Finance and Insurance Sector, Real Estate, Rent and Lease sector, retail trade sector, and healthcare and social assistance sector and the last one- manufacturing** with every variable being quarterly adjusted.

We will be using Panel Least squares estimation as the data has cross-sections and is pooled.

The main tests are Wald tests to check whether the variables are individually significant or jointly significant in the model. Also, the model could **suffer** from multicollinearity, heteroskedasticity, and correlation, hence we are doing those tests to make sure the model is adjusted before moving forward to **forecasting**.

For this model the estimated equation would be which can be used in EViews -

GDP growth rate = B_0 + B_1 Real_rent_lease + B_2 Fin_ins+ B_3 Retail trade+ B_4 healthcare_social asiss + B5manufacturing+ u (Growth rate same period previous year)

(We will use a 5% significance level for this model)

GDP growth rate = B_0 + B_1*Real_rent_lease + B_2*Fin_ins+ B_3*Retail trade+ B_4*healthcare_social asiss + B5manufacturing+ u

Is the final equation for the model as all the independent variables are significant.

As we have cross-sectional data and the data is pooled for the different states, we will be using the Panel least squares estimation. This method accounts for both cross-sectional and time series dimensions, allowing for control of heterogeneity by considering entity-specific effects. Fixed effects estimation assumes that these entity-specific effects are correlated with the independent variables, whereas random effects estimation assumes they are uncorrelated.

1.1

We can estimate the models by least squares and the estimated output tables can be seen below:– **The Midwest :**

Northeast :

Dependent Variable: GDP_GR
Method: Panel Least Squares
Date: 07/31/24 Time: 17:30
Sample: 2018Q1 2023Q4
Periods included: 24
Cross-sections included: 4
Total panel (balanced) observations: 96

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-35506.48	9279.283	-3.826424	0.0002
REAL_RENT_LEASE	4.802744	0.230946	20.79595	0.0000
FIN_INS	1.305330	0.517387	2.522929	0.0134
RETAIL_TRADE	-7.553551	1.036529	-7.287354	0.0000
HEALTHCARE_SOCIAL_ASSIST	8.811756	0.700013	12.58800	0.0000
MANUFACTURING	0.666073	0.303516	2.194521	0.0308

Root MSE	17010.83	R-squared	0.995704
Mean dependent var	607763.5	Adjusted R-squared	0.995465
S.D. dependent var	260888.8	S.E. of regression	17568.71
Akaike info criterion	22.44609	Sum squared resid	2.78E+10
Schwarz criterion	22.60636	Log likelihood	-1071.412
Hannan-Quinn criter.	22.51087	F-statistic	4171.717
Durbin-Watson stat	0.403839	Prob(F-statistic)	0.000000

\-

Dependent Variable: GDP_GR
Method: Panel Least Squares
Date: 07/31/24 Time: 17:35
Sample: 2018Q1 2023Q4
Periods included: 24
Cross-sections included: 4
Total panel (balanced) observations: 96

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-215450.1	31973.06	-6.738488	0.0000
REAL_RENT_LEASE	3.245541	0.492723	6.586950	0.0000
FIN_INS	1.574498	0.193440	8.139471	0.0000
RETAIL_TRADE	2.666366	0.952659	2.798867	0.0063
HEALTHCARE_SOCIAL_ASSIST	7.345127	0.352922	20.81235	0.0000
MANUFACTURING	-0.192245	0.039063	-4.921426	0.0000

Root MSE	26997.97	R-squared	0.997438
Mean dependent var	1003612.	Adjusted R-squared	0.997296
S.D. dependent var	536180.6	S.E. of regression	27883.39
Akaike info criterion	23.36991	Sum squared resid	7.00E+10
Schwarz criterion	23.53018	Log likelihood	-1115.756
Hannan-Quinn criter.	23.43470	F-statistic	7007.621
Durbin-Watson stat	0.241758	Prob(F-statistic)	0.000000

Both the models has an **r-squared** values of almost 0.99 which means that **99%** of the values in this model of GDP growth rate can be explained by these variables.

- The **negative coefficients** in the output of the variables which are manufacturing sector in northeast and retail trade sector in Midwest represent that a 1 unit decrease in those variables would **decrease the GDP** growth rate by the same number for the Midwest.

- For both these models, All the variables are statistically significant as all the variables have a p-value of less than 0.05.
- According to the p-values and coefficient values of variables in Midwest, The healthcare and social assistance sector along with real assts, rent and lease sector are the most significant in the model as their p-values are close to 0.0000 and their coefficient value being 8.81 and 4.80 respectively high, Hence they are the most significant sectors for GDP growth in Midwest. While the retail trade sector has an inverse relationship with the dependent variable as it's magnitude is -7.55 meaning that for every unit increase in the retail trade variable, the dependent variable (GDP growth rate) decreases by the magnitude of the coefficient, holding all other variables constant.
- For the Northeast, the same sectors healthcare and social assistance sector along with financial and insurance and real assets, rental and lease are the most significant in the Northeast where as In this case the manufacturing variable has an inverse relationship with the dependent variable and the retail trade variable is more significant than the Financial sector.

1.2

As for the Midwest, The retail trade sector and the Manufacturing sector are the least significant amongst the others in the region, Therefore we could do a Wald test to confirm their jointly significance to the dependent variable as both sectors could be related to each other as well.

The Midwest :

Wald Test:
Equation: EQ01

Test Statistic	Value	df	Probability
F-statistic	48.77118	(2, 90)	0.0000
Chi-square	97.54236	2	0.0000

Null Hypothesis: C(4)=0, C(6)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(4)	-7.553551	1.036529
C(6)	0.666073	0.303516

Restrictions are linear in coefficients.

According to the p-value, The result is that the p-values becomes much less and their joint significance is greater to the dependent variable in the region and their coefficient is significantly different from zero.

2

Hypothesis Testing

Now we will do hypothesis testing for the manufacturing variable which can be done through the Wald Test to determine whether the manufacturing variable affects the dependent variable GDP growth rate significantly or not in the **Northeast region** as magnitude is negative.

- For that the Null Hypothesis will be H_0 – The retail trade variable affects the dependent variable significantly.
- Alternative hypothesis H_1 – The variable do not affect the GDP growth rate significantly.

Wald Test:
Equation: EQ01

Test Statistic	Value	df	Probability
t-statistic	-4.921426	90	0.0000
F-statistic	24.22043	(1, 90)	0.0000
Chi-square	24.22043	1	0.0000

Null Hypothesis: C(6)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(6)	-0.192245	0.039063

Restrictions are linear in coefficients.

The p-value being less than 0.05, we can say that we can reject the null hypothesis and the alternative hypothesis is true and the retail trade variable is contributing significantly to the GDP growth in Northeast region as well with the other variables/sectors.

2.1

Cross-sectional Dependence

Cross-sectional dependence in estimation model can cause inaccurate estimation output and results. An cross-section dependence test is necessary to know if there is cross-sectional dependency in the models or not as ignoring cross-sectional dependence can lead to biased and inefficient parameter estimates.

The output is given below:

The Midwest :

Residual Cross-Section Dependence Test
Null hypothesis: No cross-section dependence (correlation) in residuals
Equation: EQ01
Periods included: 24
Cross-sections included: 4
Total panel observations: 96
Note: non-zero cross-section means detected in data
Cross-section means were removed during computation of correlations

Test	Statistic	d.f.	Prob.
Breusch-Pagan LM	55.63507	6	0.0000
Pesaran scaled LM	14.32841		0.0000
Pesaran CD	1.798900		0.0720

The Northeast :

Residual Cross-Section Dependence Test
Null hypothesis: No cross-section dependence (correlation) in residuals
Equation: EQ01
Periods included: 24
Cross-sections included: 4
Total panel observations: 96
Note: non-zero cross-section means detected in data
Cross-section means were removed during computation of correlations

Test	Statistic	d.f.	Prob.
Breusch-Pagan LM	46.31185	6	0.0000
Pesaran scaled LM	11.63703		0.0000
Pesaran CD	-0.652350		0.5142

As the P-value is less than alpha (0.05) for the Breusch-Pagan LM and Pesaran scaled LM test we could reject the null hypothesis and state that there is correlation in the residual terms in both the regions but according to the Pesaran CD Test there is no correlation in the residuals in both the models as their p-value is greater than 0.05 i.e significance level.

2.2

Adjusting for heteroskedasticity with White cross-section estimation for both the regions as heteroskedasticity reduces the accuracy of the estimations for the models for both the regions :

The Midwest :

Dependent Variable: GDP_GR
Method: Panel Least Squares
Date: 07/31/24 Time: 19:34
Sample: 2018Q1 2023Q4
Periods included: 24
Cross-sections included: 4
Total panel (balanced) observations: 96
White cross-section standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-35506.48	7511.207	-4.727133	0.0000
REAL_RENT_LEASE	4.802744	0.166836	28.78726	0.0000
FIN_INS	1.305330	0.472354	2.763458	0.0069
RETAIL_TRADE	-7.553551	1.094498	-6.901387	0.0000
HEALTHCARE_SOCIAL_ASSIST	8.811756	0.721076	12.22028	0.0000
MANUFACTURING	0.666073	0.326782	2.038279	0.0445

Root MSE	17010.83	R-squared	0.995704
Mean dependent var	607763.5	Adjusted R-squared	0.995465
S.D. dependent var	260888.8	S.E. of regression	17568.71
Akaike info criterion	22.44609	Sum squared resid	2.78E+10
Schwarz criterion	22.60636	Log likelihood	-1071.412
Hannan-Quinn criter.	22.51087	F-statistic	4171.717
Durbin-Watson stat	0.403839	Prob(F-statistic)	0.000000

Northeast :

Dependent Variable: GDP_GR
Method: Panel Least Squares
Date: 07/31/24 Time: 19:38
Sample: 2018Q1 2023Q4
Periods included: 24
Cross-sections included: 4
Total panel (balanced) observations: 96
White cross-section standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-215450.1	35260.56	-6.110228	0.0000
REAL_RENT_LEASE	3.245541	0.501750	6.468446	0.0000
FIN_INS	1.574498	0.125422	12.55365	0.0000
RETAIL_TRADE	2.666366	0.971492	2.744609	0.0073
HEALTHCARE_SOCIAL_ASSIST	7.345127	0.446169	16.46267	0.0000
MANUFACTURING	-0.192245	0.035679	-5.388107	0.0000

Root MSE	26997.97	R-squared	0.997438
Mean dependent var	1003612.	Adjusted R-squared	0.997296
S.D. dependent var	536180.6	S.E. of regression	27883.39
Akaike info criterion	23.36991	Sum squared resid	7.00E+10
Schwarz criterion	23.53018	Log likelihood	-1115.756
Hannan-Quinn criter.	23.43470	F-statistic	7007.621
Durbin-Watson stat	0.241758	Prob(F-statistic)	0.000000

Here the standard errors and the p-values are adjusted now and are valid because we have adjusted those values with the White cross-section test which contains consistent standard errors and covariance.

2.3

Serial correlations in models can cause unreliable hypothesis test results in regression models, So there’s a need to check if the models have a serial correlation or not which is shown below with the estimation adjustment with Robust Least squares estimation which adjusts the model for correlation :

The Midwest :

Dependent Variable: GDP_GR
Method: Robust Least Squares
Date: 07/31/24 Time: 19:57
Sample: 2018Q1 2023Q4
Included observations: 96
Method: M-estimation
M settings: weight=Bisquare, tuning=4.685, scale=MAD (median centered)
Huber Type I Standard Errors & Covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-25219.45	7999.414	-3.152663	0.0016
REAL_RENT_LEASE	4.690200	0.199092	23.55792	0.0000
FIN_INS	1.387976	0.446025	3.111879	0.0019
RETAIL_TRADE	-6.851704	0.893563	-7.667849	0.0000
HEALTHCARE_SOCIAL_ASSIST	8.194864	0.603462	13.57976	0.0000
MANUFACTURING	0.651886	0.261653	2.491412	0.0127

Robust Statistics

R-squared	0.848751	Adjusted R-squared	0.840348
Rw-squared	0.997269	Adjust Rw-squared	0.997269
Akaike info criterion	104.3222	Schwarz criterion	122.9935
Deviance	2.05E+10	Scale	14628.05
Rn-squared statistic	27164.28	Prob(Rn-squared stat.)	0.000000

Non-robust Statistics

Mean dependent var	607763.5	S.D. dependent var	260888.8
S.E. of regression	18393.21	Sum squared resid	3.04E+10

Northeast :

Dependent Variable: GDP_GR
Method: Robust Least Squares
Date: 07/31/24 Time: 19:44
Sample: 2018Q1 2023Q4
Included observations: 96
Method: M-estimation
M settings: weight=Bisquare, tuning=4.685, scale=MAD (median centered)
Huber Type I Standard Errors & Covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-77522.55	21269.21	-3.644825	0.0003
REAL_RENT_LEASE	2.132338	0.327771	6.505581	0.0000
FIN_INS	1.662087	0.128681	12.91637	0.0000
RETAIL_TRADE	4.860325	0.633731	7.669387	0.0000
HEALTHCARE_SOCIAL_ASSIST	5.053346	0.234772	21.52452	0.0000
MANUFACTURING	-0.106606	0.025986	-4.102537	0.0000

Robust Statistics

R-squared	0.785412	Adjusted R-squared	0.773490
Rw-squared	0.999205	Adjust Rw-squared	0.999205
Akaike info criterion	150.0745	Schwarz criterion	169.3722
Deviance	4.38E+10	Scale	17560.72
Rn-squared statistic	80335.98	Prob(Rn-squared stat.)	0.000000

Non-robust Statistics

Mean dependent var	1003612.	S.D. dependent var	536180.6
S.E. of regression	36541.49	Sum squared resid	1.20E+11

Both the models for both the regions have been accounted for adjustment to correlation and now we can proceed further.

2.4

Auto Regressive models

Now to construct an AR model that helps to adjust a model to account for the dependence of a variable on its own past values we will use a correlogram to check which AR terms could be added to the model.

The Midwest :

Date: 07/31/24 Time: 20:12
Sample: 2018Q1 2023Q4
Included observations: 96

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.728	0.728	52.436	0.000
		2	0.582	0.111	86.301	0.000
		3	0.466	0.020	108.28	0.000
		4	0.306	-0.143	117.84	0.000
		5	0.212	0.000	122.46	0.000
		6	0.105	-0.077	123.62	0.000
		7	0.063	0.059	124.04	0.000
		8	0.012	-0.044	124.06	0.000
		9	0.019	0.090	124.10	0.000
		10	0.087	0.139	124.92	0.000
		11	0.055	-0.104	125.26	0.000
		12	-0.008	-0.169	125.26	0.000
		13	0.018	0.097	125.30	0.000
		14	-0.025	-0.061	125.37	0.000
		15	-0.069	-0.051	125.92	0.000
		16	-0.103	-0.048	127.16	0.000

*Probabilities may not be valid for this equation specification.

Northeast :

Date: 07/31/24 Time: 20:10
Sample: 2018Q1 2023Q4
Included observations: 96

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.876	0.876	75.971	0.000
		2	0.746	-0.090	131.69	0.000
		3	0.616	-0.076	170.05	0.000
		4	0.496	-0.036	195.17	0.000
		5	0.363	-0.135	208.83	0.000
		6	0.265	0.055	216.16	0.000
		7	0.169	-0.074	219.19	0.000
		8	0.087	-0.029	220.00	0.000
		9	0.003	-0.081	220.00	0.000
		10	-0.062	-0.020	220.43	0.000
		11	-0.116	-0.014	221.92	0.000
		12	-0.127	0.106	223.73	0.000
		13	-0.138	-0.044	225.88	0.000
		14	-0.117	0.086	227.47	0.000
		15	-0.095	-0.008	228.52	0.000
		16	-0.088	-0.099	229.43	0.000

*Probabilities may not be valid for this equation specification.

- It is observed that AR(1) is more prominent in Midwest and AR (1) is more prominent in Northeast than others, We will add these terms to the model.

Therefore with the ARMA Maximum Likelihood (OPG – BHHH) method, this can be done and the output Table is given below:-

The Midwest :

Dependent Variable: GDP_GR
Method: Panel Least Squares
Date: 07/31/24 Time: 20:13
Sample (adjusted): 2018Q2 2023Q4
Periods included: 23
Cross-sections included: 4
Total panel (balanced) observations: 92
Convergence achieved after 52 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-62999.31	24298.53	-2.592721	0.0112
REAL_RENT_LEASE	4.587402	0.475727	9.642938	0.0000
FIN_INS	0.824796	0.220992	3.732236	0.0003
RETAIL_TRADE	1.301440	0.431827	3.013797	0.0034
HEALTHCARE_SOCIAL_ASSIST	2.861159	0.385328	7.425260	0.0000
MANUFACTURING	2.128527	0.238103	8.939519	0.0000
AR(1)	0.954918	0.014080	67.82015	0.0000

Root MSE	4069.302	R-squared	0.999756
Mean dependent var	610588.8	Adjusted R-squared	0.999739
S.D. dependent var	261807.2	S.E. of regression	4233.547
Akaike info criterion	19.61250	Sum squared resid	1.52E+09
Schwarz criterion	19.80438	Log likelihood	-895.1752
Hannan-Quinn criter.	19.68995	F-statistic	57988.06
Durbin-Watson stat	1.433934	Prob(F-statistic)	0.000000

Inverted AR Roots	.95

Dependent Variable: GDP_GR
Method: Panel Least Squares
Date: 07/31/24 Time: 20:13
Sample (adjusted): 2018Q2 2023Q4
Periods included: 23
Cross-sections included: 4
Total panel (balanced) observations: 92
Convergence achieved after 21 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	462766.9	598358.1	0.773395	0.4414
REAL_RENT_LEASE	2.107356	0.352981	5.970165	0.0000
FIN_INS	1.283141	0.224692	5.710666	0.0000
RETAIL_TRADE	0.827853	0.622496	1.329892	0.1871
HEALTHCARE_SOCIAL_ASSIST	5.785196	0.490094	11.80427	0.0000
MANUFACTURING	0.133274	0.038913	3.424914	0.0009
AR(1)	0.992202	0.010173	97.53240	0.0000

Root MSE	9018.205	R-squared	0.999717
Mean dependent var	1007985.	Adjusted R-squared	0.999697
S.D. dependent var	538891.1	S.E. of regression	9382.197
Akaike info criterion	21.20405	Sum squared resid	7.48E+09
Schwarz criterion	21.39593	Log likelihood	-968.3864
Hannan-Quinn criter.	21.28149	F-statistic	50021.89
Durbin-Watson stat	1.155568	Prob(F-statistic)	0.000000

Inverted AR Roots	.99

Northeast :

The **addition** of AR terms causes r-square value to increase slightly in both the models and which can significantly enhance the model by incorporating past values of the dependent variable as predictors for help in forecasting.

Also the Finance and insurance sector becomes more significant in the Northeast after the addition of the AR term and retail trade variable becomes insignificant after the addition of AR(1).

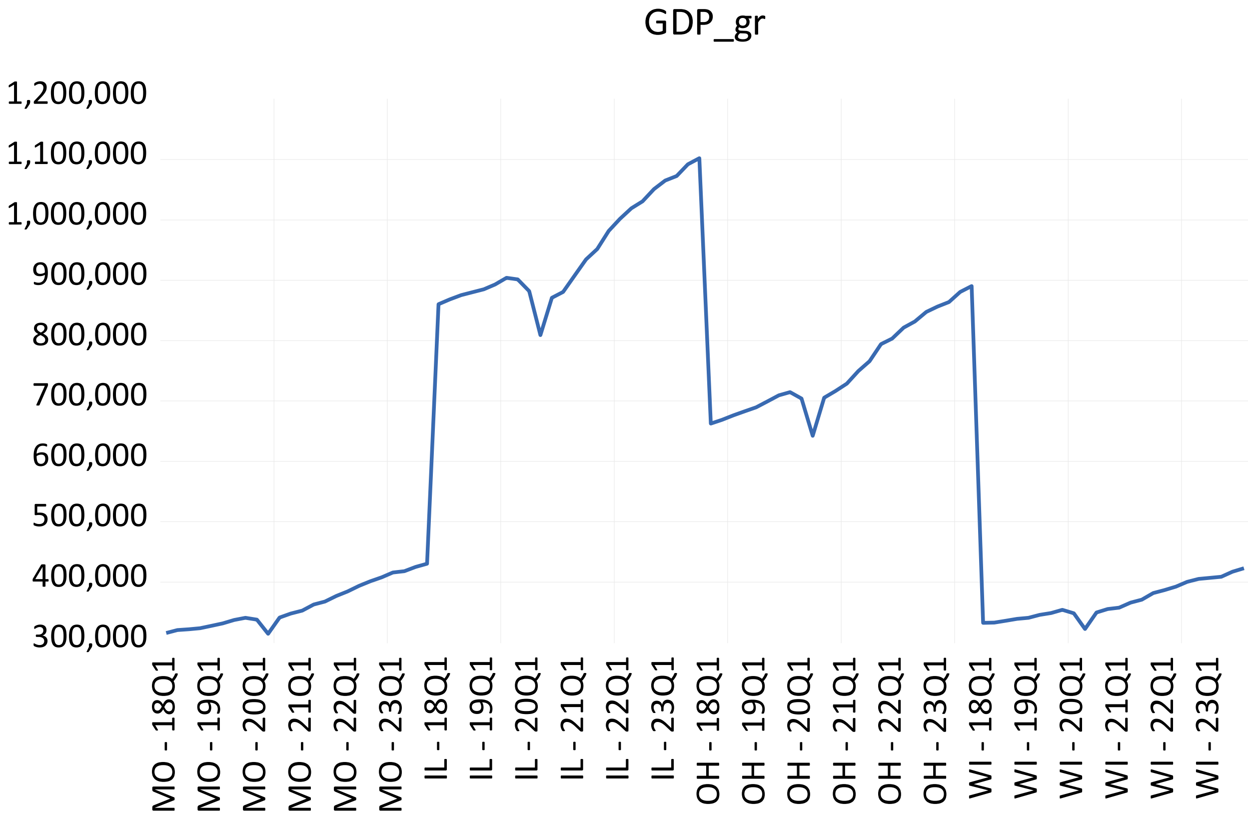
3.1

Testing stationary on non-stationary variables:-

Whether the dependent variable is stationary or not is important to know in a model as non–stationary variables can lead to misleading conclusions in the model and improper forecasting. Hence we will observe whether the dependent variable in our model which is the GDP growth rate is stationary or not and how to adjust it.

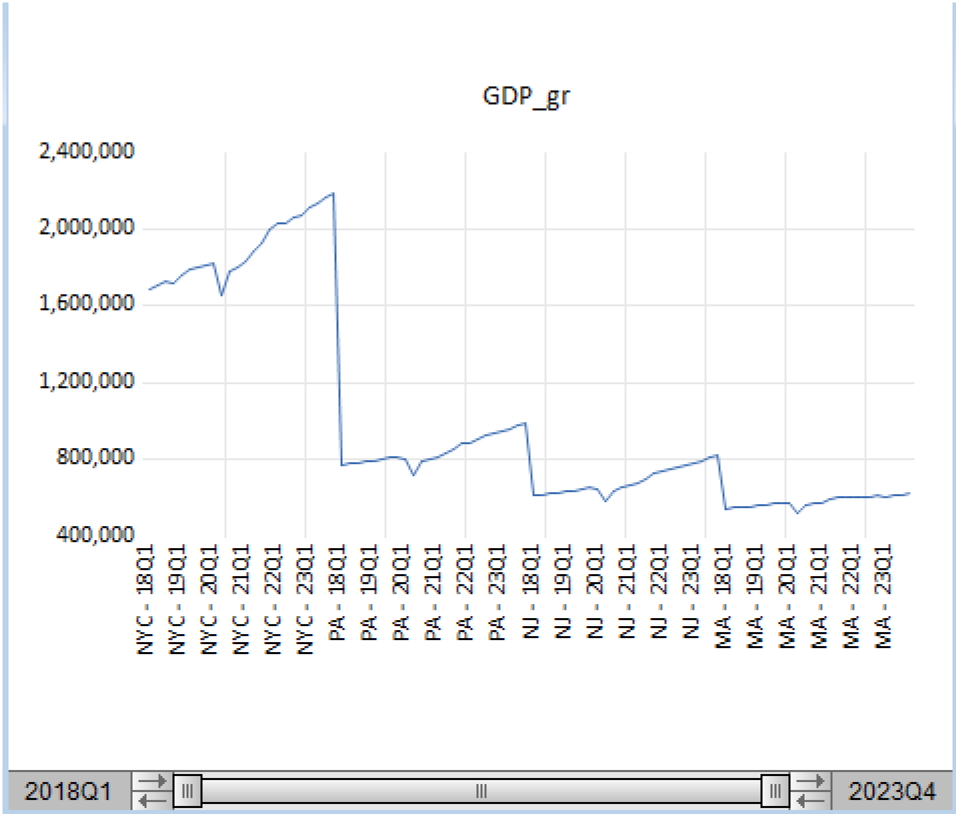
A graph of GDP growth rate is given below for both the regions –

1. Midwest –



The GDP growth rate ratio appears to be stationary in the long run in this case as it often revolves around the mean after a few years till now.

2. Northeast :



The GDP growth rate appears to be non-stationary as the data in the model is limited but it the GDP growth rate as a whole is a stationary variable if instead of quarterly data, we use the yearly frequency in the data.

3.2

Unit root tests help determine whether a time series is stationary or non-stationary. Stationarity is an important assumption in many time series models because it ensures that the statistical properties of the series, such as mean and variance, remain constant over time.

Hence we will now check whether the GDP growth rate contains a unit root or not for both the regions-

As the real estate, rent and lease variable affects the GDP growth rate the most in Midwest region we will check whether the variable has a unit root or not and will be using the ADF test.

Panel unit root test: Summary
Series: REAL_RENT_LEASE
Date: 07/31/24 Time: 20:21
Sample: 2018Q1 2023Q4
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	0.93645	0.8255	4	88
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	1.32120	0.9068	4	88
ADF - Fisher Chi-square	4.22096	0.8367	4	88
PP - Fisher Chi-square	3.32334	0.9125	4	92

** Probabilities for Fisher tests are computed using an asymptotic Chi-square distribution. All other tests assume asymptotic normality.

The null hypothesis for the unit root test states that the variable has a unit root and observing the p-value of the test we say that the null hypothesis can be rejected and the alternative hypothesis is true that the real asset, rental and lease variable does not contain a unit root.

Northeast-

For the Northeast region, we will use the ADF test with the variable finance and insurance to check if it contains a unit root or not.

Panel unit root test: Summary

Series: FIN_INS

Date: 07/31/24 Time: 20:23

Sample: 2018Q1 2023Q4
Exogenous variables: Individual effects

Automatic selection of maximum lags
Automatic lag length selection based on SIC: 0

Newey-West automatic bandwidth selection and Bartlett kernel

Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-0.86328	0.1940	4	92
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-0.96854	0.1664	4	92
ADF - Fisher Chi-square	14.0696	0.0800	4	92
PP - Fisher Chi-square	14.3926	0.0721	4	92

** Probabilities for Fisher tests are computed using an asymptotic Chi

-square distribution. All other tests assume asymptotic normality.

As the p-value is greater than alpha(0.05) we can reject the null hypothesis and the alternative hypothesis is true that the financial and insurance variable does not contain a unit root.

4.0

Forecasting

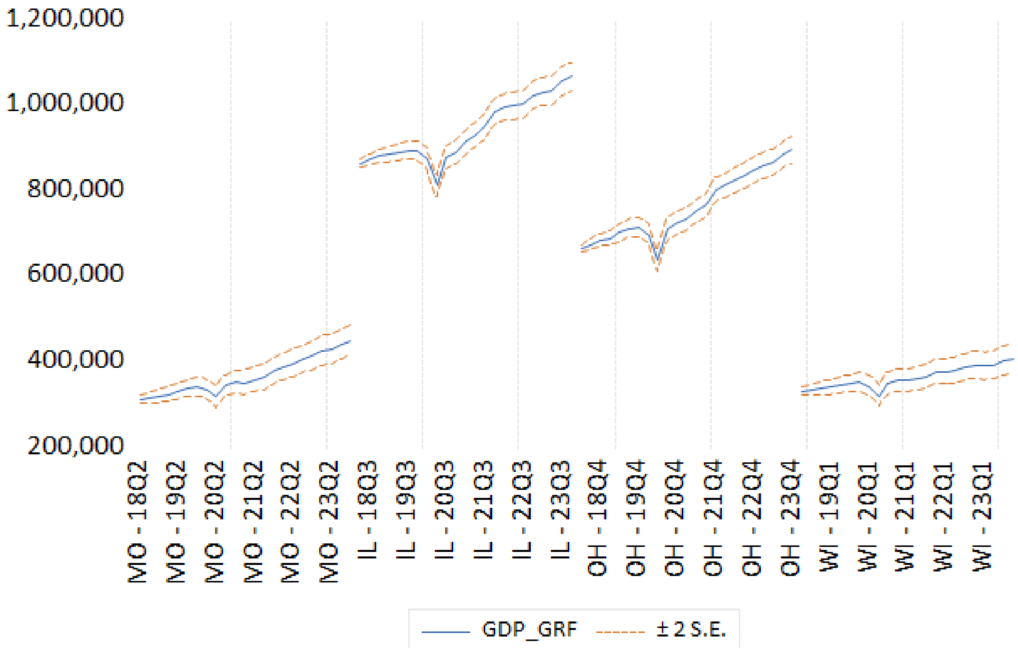
As the model is appropriate now we can go ahead with the forecasting, Forecasting is typically associated with predicting future values based on historical data.

We will compare the dynamic and static forecasts for both the regions to check which forecast will be appropriate for the model.

Forecasting serves as a means to evaluate the performance of the model. By comparing the model's forecasts to the actual observed values within the sample period, we can assess how well the model captures the underlying patterns in the data and compare the dynamic and static forecasts which could be a better one for the model.

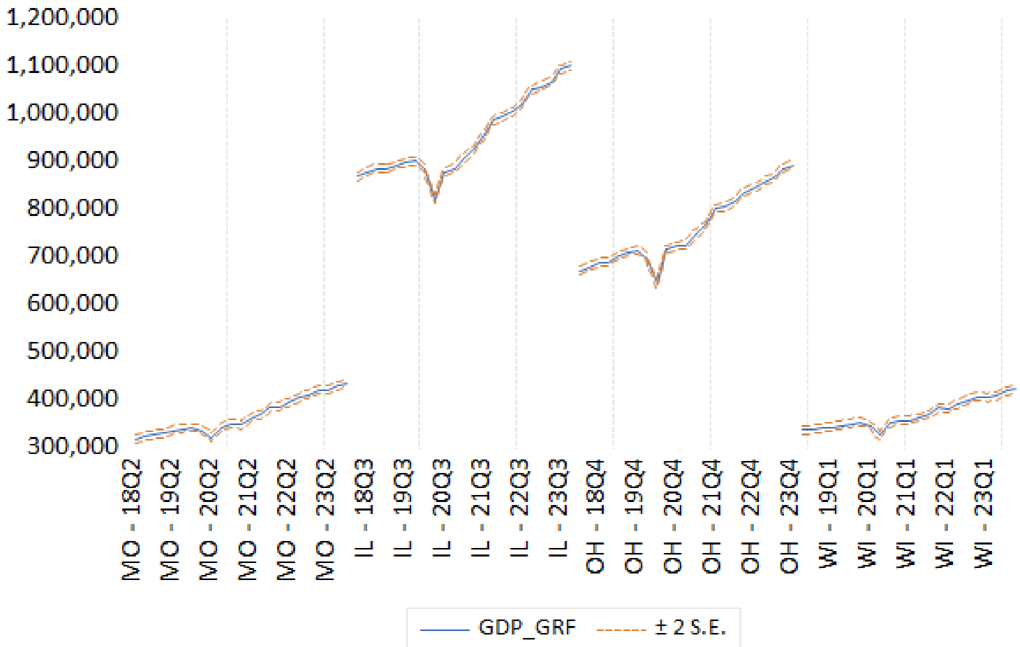
Given below is the dynamic and static forecast of all of the sample data for the **Midwest** region–

Dynamic Forecast -



Forecast: GDP_GRF	
Actual: GDP_GR	
Forecast sample: 2018Q1 2023Q4	
Adjusted sample: 2018Q2 2023Q4	
Included observations: 92	
Root Mean Squared Error	10831.28
Mean Absolute Error	8044.265
Mean Abs. Percent Error	1.374419
Theil Inequality Coef.	0.008156
Bias Proportion	0.024475
Variance Proportion	0.073539
Covariance Proportion	0.901987
Theil U2 Coefficient	0.610608
Symmetric MAPE	1.369788

Static Forecast –



Forecast: GDP_GRF	
Actual: GDP_GR	
Forecast sample: 2018Q1 2023Q4	
Adjusted sample: 2018Q2 2023Q4	
Included observations: 92	
Root Mean Squared Error	4069.302
Mean Absolute Error	2935.808
Mean Abs. Percent Error	0.513492
Theil Inequality Coef.	0.003066
Bias Proportion	0.000000
Variance Proportion	0.037576
Covariance Proportion	0.962424
Theil U2 Coefficient	0.230778
Symmetric MAPE	0.513307

When we compare the Root Mean Squared Error, Mean Absolute error, and Mean Abs. Percentage error of both the forecasts which are all measures of forecast accuracy.

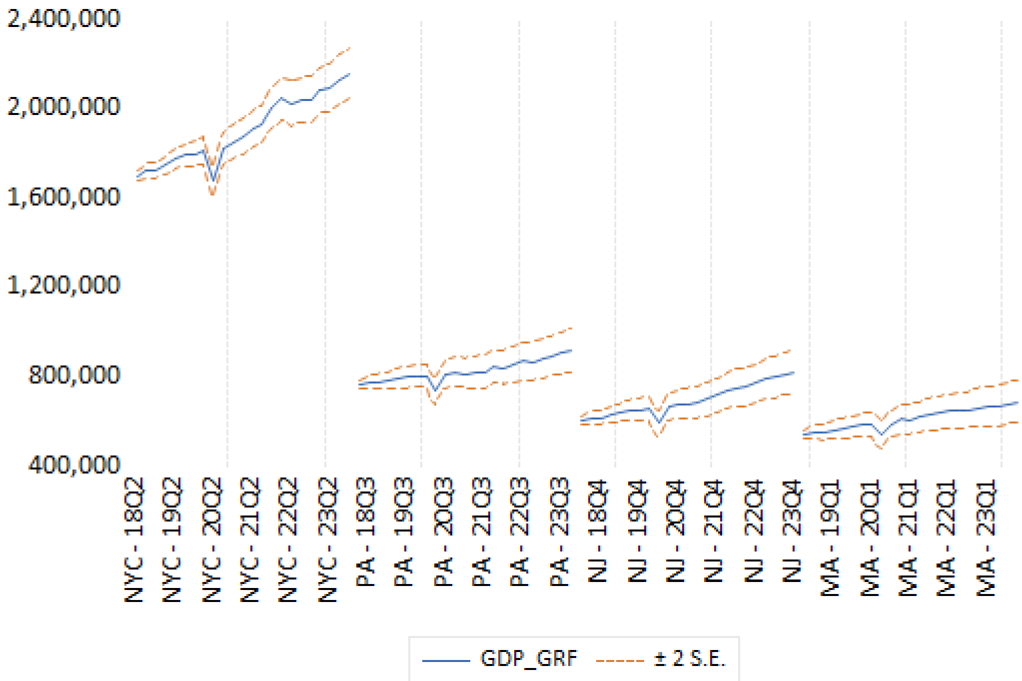
The RMSE focuses on large errors, MAE gives equal weight to all errors, and MAPE provides a percentage measure of error relative to the actual values.

We notice that in the Static forecast, all the values of the RMSE, MAE, and MAPE are smaller than those of the Dynamic forecast.

Therefore it can be said that the **Dynamic forecast** is better in this context for the Midwest region.

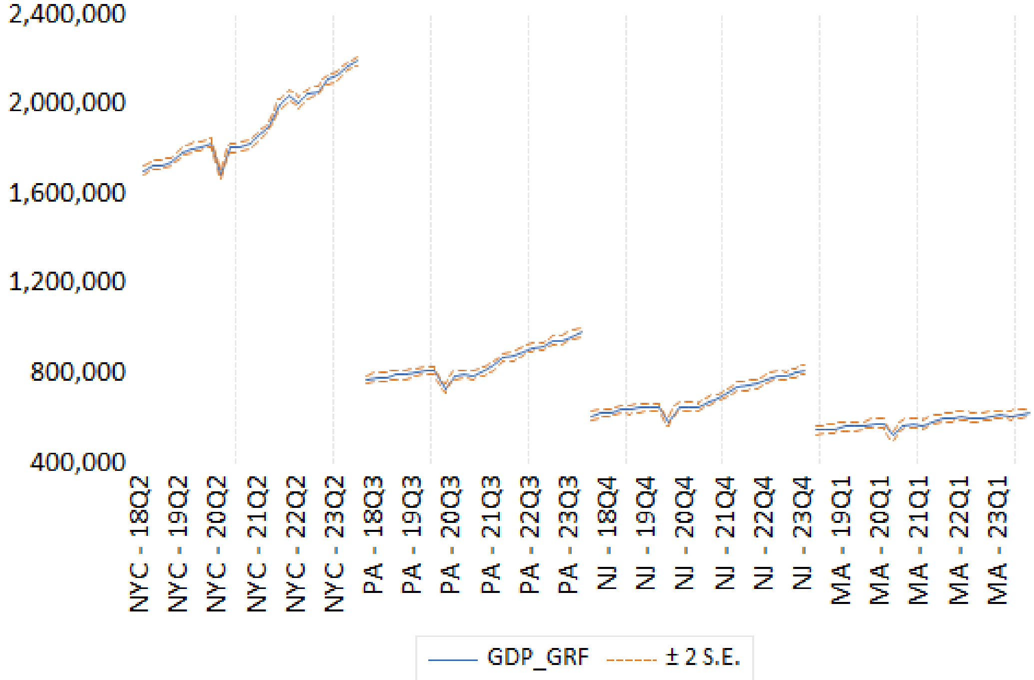
Similarly, Dynamic Forecast

for **Northeast**-



Forecast: GDP_GRF	
Actual: GDP_GR	
Forecast sample: 2018Q1 2023Q4	
Adjusted sample: 2018Q2 2023Q4	
Included observations: 92	
Root Mean Squared Error	30679.43
Mean Absolute Error	22981.18
Mean Abs. Percent Error	2.857801
Theil Inequality Coef.	0.013414
Bias Proportion	0.065681
Variance Proportion	0.046330
Covariance Proportion	0.887988
Theil U2 Coefficient	1.358395
Symmetric MAPE	2.807471

Static Forecast-



Forecast: GDP_GRF	
Actual: GDP_GR	
Forecast sample: 2018Q1 2023Q4	
Adjusted sample: 2018Q2 2023Q4	
Included observations: 92	
Root Mean Squared Error	9018.205
Mean Absolute Error	6386.692
Mean Abs. Percent Error	0.683095
Theil Inequality Coef.	0.003951
Bias Proportion	0.000000
Variance Proportion	0.021629
Covariance Proportion	0.978371
Theil U2 Coefficient	0.305226
Symmetric MAPE	0.682390

Similarly, In this case, The static forecast it more appropriate than the dynamic forecast as it has smaller values for the RMSE, MAE and MAPE. Hence it is more appropriate for the Northeast Region.

Findings and conclusions –

- In our study all the variables which affected the region significantly were significant such as the real assets, rental and lease variable for Midwest was the most contributing sector for the Midwest region where as for the Northeast it was Finance and insurance and healthcare and social assistance.
- The model suffered from heteroskedasticity and was adjusted accordingly.
- For our study, the static forecast was suitable for the Midwest region as well as the static forecast was more appropriate for the Northeast region.
- The forecasts were accurate and worked for both regions.
- The model explains how the sectors affect the different regions in the U.S. and what sectors contribute more in a particular region in the U.S. This helps the analysts and policy makers to take steps and focus more on what factors are the regions affected by the most and steps to improve other sectors efficiently as well. One example could be the steps to increase security in the Midwest as more people would feel safe and move in the cities instead of moving out and become more populated once more as it was before and thriving in the financial sector.
- The GDP growth rate of New York is significantly higher than the other states in the Northeast region which could be due to the popularity of the place and high income as well as living standards.
- For Midwest, Illinois has the highest GDP contribution as Chicago would be the place to contribute the most due to the same reasons as New York.

To improve the economy by focusing on different sectors in specific regions, targeted strategies can leverage the unique strengths of each area. In the Midwest, enhancing the real estate sector through affordable housing projects, commercial infrastructure development, and regulatory support can stimulate construction, attract businesses, and create jobs, driving GDP growth. Additionally, investing in agricultural and manufacturing advancements, improving supply chain logistics, and providing workforce training can boost productivity and competitiveness. In the Northeast, focusing on the financial sector, through regulatory enhancements and fostering financial innovation, can attract investment and enhance economic stability, contributing significantly to GDP growth. These region-specific strategies can lead to more robust and sustainable economic development across the U.S. where as focusing on other sectors more than the most contributing sector to improve the economy more in the particular region which will make the region more economically stable.

In all the model has been adjusted and is able to explain the variables included in it with the forecasts being accurate.

References –

- (https://fred.stlouisfed.org/)