Warith Scientific Journal

Analysis of Factors Influencing Sustainable Development in Iraq Using Statistical Modeling: An Analysis of the Unemployment Rate for the Period 2010 to 2020

تحليل العوامل المؤثرة في تحقيق التنمية المستدامة في العراق باستخدام النمذجة الإحصائية: تحليل لمعدل البطالة للفترة من 2010 إلى 2020

DR. Saif Hosam Raheem[1]

د. سيف حسام رحيم[1]

University of Al-Qadisiyah, Collage of Administration and Economics, Department of statistics

جامعة القادسية، كلية الإدارة والأقتصاد، قسم الإحصاء

saif.hosam@qu.edu.iq

DR. Fadel Hamid Hadi Alhusseini[2]

د. فاضل حميد هادي الحسيني[2]

University of Al-Qadisiyah, Collage of Administration and Economics, Department of statistics

جامعة القادسية، كلية الإدارة والأقتصاد، قسم الإحصاء

fadel.alhsiny@qu.edu.iq

المستخلص:

تهدف هذه الدراسة إلى تحليل العوامل المؤثرة في التنمية المستدامة في العراق، مع التركيز بشكل خاص على معدل البطالة، باستخدام تقنيات النمذجة الإحصائية. تغطي فترة الدراسة من 2010 إلى 2020، وهي فترة حاسمة في تاريخ العراق حيث شهدت تغييرات اقتصادية واجتماعية وسياسية هامة أثرت على الوضع التنموي في البلاد. تستخدم الدراسة نموذج الانحدار التوقعي (Expectile Regression) لفحص العلاقة بين العوامل الاقتصادية والاجتماعية والسياسية، بما في ذلك البطالة، وأثرها على التنمية المستدامة في العراق. تظهر النتائج أهمية هذه العوامل في تشكيل المشهد الاجتماعي والاقتصادي للعراق، مما يوفر رؤى قيمة لصانعي السياسات والمختصين في التنمية. من خلال استخدام الأساليب الإحصائية، يقدم هذا البحث إطارًا قويًا لفهم ديناميكيات التنمية المستدامة في العراق ويحدد المجالات الرئيسية التي تحتاج إلى تحسين.

الكلمات المفتاحية: التنمية المستدامة، معدل البطالة، الانحدار المتوقع

Abstract:

This study aims to analyze the factors influencing sustainable development in Iraq, with a specific focus on the unemployment rate, using statistical modeling techniques. The period under investigation spans from 2010 to 2020, a critical period in Iraq's history that witnessed significant economic, social, and political changes affecting the country's development status. The study employs Expectile Regression to examine the relationship between economic, social, and political factors, including unemployment, and their impact on sustainable development in the country. The results highlight the significance of these factors in shaping the socio-economic landscape of Iraq, providing valuable insights for policymakers and development professionals. By utilizing statistical methods, this research offers a robust framework for understanding the dynamics of sustainable development in Iraq and identifies key areas for improvement.

Keywords: Sustainable development, unemployment rate, Expectile Regression.

1. Introduction

Sustainable development has become a central objective for countries seeking to balance economic growth, social equity, and environmental protection. In Iraq, the pursuit of sustainable development faces complex challenges, particularly in the presence of economic instability, political unrest, and rising unemployment. The unemployment rate is not only a key indicator of

economic performance but also a critical factor influencing long-term development outcomes, including poverty, inequality, and social cohesion.

Analyzing the factors that affect sustainable development in Iraq requires a methodological framework that can capture complex and asymmetric relationships. Traditional regression models often fall short in dealing with heteroscedasticity, skewed distributions, and sensitivity to outliers. In this context, Expectile Regression, first introduced by Newey and Powell (1987), offers a robust alternative. It estimates conditional expectiles instead of means or quantiles, allowing for a richer characterization of the distribution of the dependent variable.

Expectile regression is particularly useful when the effects of explanatory variables vary across the distribution of the outcome, as seen in labor market indicators such as unemployment. The model's ability to emphasize different parts of the conditional distribution makes it suitable for examining both central trends and extreme values (Waltrup et al., 2015; Sobotka et al., 2021). Moreover, recent developments have shown the flexibility of expectile regression in capturing asymmetric and heavy-tailed error structures, making it advantageous over quantile regression in some economic contexts (Yu and Moyeed, 2001; Schnabel and Eilers, 2009).

This study applies expectile regression to assess the impact of economic, political, and social factors on sustainable development in Iraq from 2010 to 2020. By focusing on unemployment as a central variable, the research aims to provide a more detailed statistical representation of development dynamics and to inform effective policy responses.

2. Methodology

This study aims to investigate the relationship between key economic, social, and political factors and sustainable development in Iraq, with a specific focus on the unemployment rate. The analysis is conducted using data from 2010 to 2020, which includes annual observations on selected macroeconomic and social indicators.

To capture the asymmetric influence of covariates across different levels of the response variable, we employ Expectile Regression, a method introduced by Newey and Powell (1987). Unlike traditional least squares regression, which estimates the conditional mean, expectile regression models various points (expectiles) of the conditional distribution, offering a richer understanding of variable impacts.

The expectile regression model is formulated as:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\tau}} = \arg\min \sum_{i=1}^{n} \boldsymbol{\rho}_{\boldsymbol{\tau}} \left(\boldsymbol{y}_i - \boldsymbol{x}_i^{\boldsymbol{T}} \boldsymbol{\beta}\right)$$

where $\rho_\tau(u) = |\tau - \mathrm{I}(\mathrm{u} < 0)|u^2$, and $\tau \in (0,1)$ specifies the expectile level. The estimation is carried out using the Iterative Weighted Least Squares (IWLS) algorithm as proposed by Newey and Powell (1987) and extended in later works (Sobotka et al., 2021; Waltrup et al., 2015).

Several expectile levels are considered (e.g., τ=0.25,0.5,0.75) to assess whether explanatory variables exert stronger effects in the lower or upper tails of the response distribution, a key advantage of expectile modeling (Yu and Jones, 1998).

Before modeling, all time series variables are tested for stationarity using the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979). Multicollinearity is assessed through the Variance Inflation Factor (VIF). Model performance is evaluated using Mean Squared Error (MSE) and visual diagnostics.

All analyses are performed using R statistical software, particularly the expectreg package, which facilitates expectile regression estimation and diagnostics.

3. Simulation Study

To evaluate the performance of Expectile Regression in capturing the asymmetric effects of predictors on the outcome variable, a simulation study is conducted. The objective is to assess the model's robustness and accuracy under different distributional settings, particularly in the presence of heteroscedasticity and skewness, which are common in economic data.
We simulate data using the following model structure:

$$\boldsymbol{y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i}\,, \qquad \mathbf{i = 1, \dots, n}$$

Where:

$$\boldsymbol{x_{i1}, x_{i2} \sim N(0,1), \beta_1 = 1.5, \beta_2 = -2}\,, \boldsymbol{\varepsilon_i} \sim \textbf{Skewed Normal } (\boldsymbol{\xi = 0, \omega = 1, \alpha = 5})$$

Three scenarios are considered:

1. Scenario A: Homoscedastic, symmetric error ($\varepsilon_i \sim N(0,1)$)
2. Scenario B: Heteroscedastic error ($\varepsilon_i \sim N(0, x_{i1}^2)$)
3. Scenario C: Skewed error using the skew-normal distribution

For each scenario:

- Sample size n=500 .
- Number of simulations: 100 .
- Expectile levels $\tau = 0.25, 0.5, 0.75$.

To compare the performance of Expectile Regression, we use the following metrics:

Mean Squared Error (MSE) between estimated and true values

$$\textbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

Bias in parameter estimation

$$Bias(\hat{\beta}) = E[\hat{\beta}] - \beta$$

Where:

- $\hat{\beta}$ is the estimated parameter.
- β is the true parameter value.
- The expectation E[·]is taken over repeated samples.

Table 1: Mean Squared Error (MSE) under different scenarios

Scenario	τ = 0.25	τ = 0.50	τ = 0.75
A	0.145	0.122	0.150
B	0.160	0.135	0.171
C	0.181	0.142	0.198

Table 1 shows the Mean Squared Error (MSE) of expectile regression estimates across three scenarios and three expectile levels (τ = 0.25, 0.50, 0.75). The results indicate that the lowest MSE values occur consistently at τ = 0.50, corresponding to the central tendency of the distribution. Scenario A, which assumes homoscedastic and symmetric errors, yields the smallest errors across all τ levels. In contrast, Scenario C, which includes skewed errors, produces the highest MSE values, especially at τ = 0.75, reflecting the model's sensitivity to distributional asymmetry at the upper tail.

Overall, the results demonstrate that expectile regression performs best under ideal conditions (Scenario A) and remains relatively robust under moderate heteroscedasticity (Scenario

B). However, its accuracy decreases under strong skewness (Scenario C), particularly at extreme expectile levels.

Table 2: Bias in estimation of β_1 and β_2 (averaged over 100 runs)

Scenario	β_1 Bias ($\tau = 0.5$)	β_2 Bias ($\tau = 0.5$)
A	**0.012**	**-0.019**
B	**0.025**	**-0.027**
C	**0.040**	**-0.052**

Table 2 presents the average bias in the estimation of the parameters β_1 and β_2 at the central expectile level τ=0.50 across 100 simulation runs. The bias is calculated as the average difference between the estimated coefficients and their true values.

In Scenario A (symmetric and homoscedastic errors), the biases for both parameters are minimal, indicating accurate estimation under ideal conditions. Scenario B introduces heteroscedasticity, which leads to a slight increase in bias for both β_1 and β_2, but the estimates remain reasonably close to the true values. Scenario C, which involves skewed errors, shows the highest bias levels, particularly for β_2, suggesting that expectile regression is more affected by distributional asymmetry in estimating parameters associated with stronger negative effects.

These findings support the reliability of expectile regression under moderate departures from classical assumptions but also highlight some sensitivity to skewness, especially in tail-related parameter estimation.

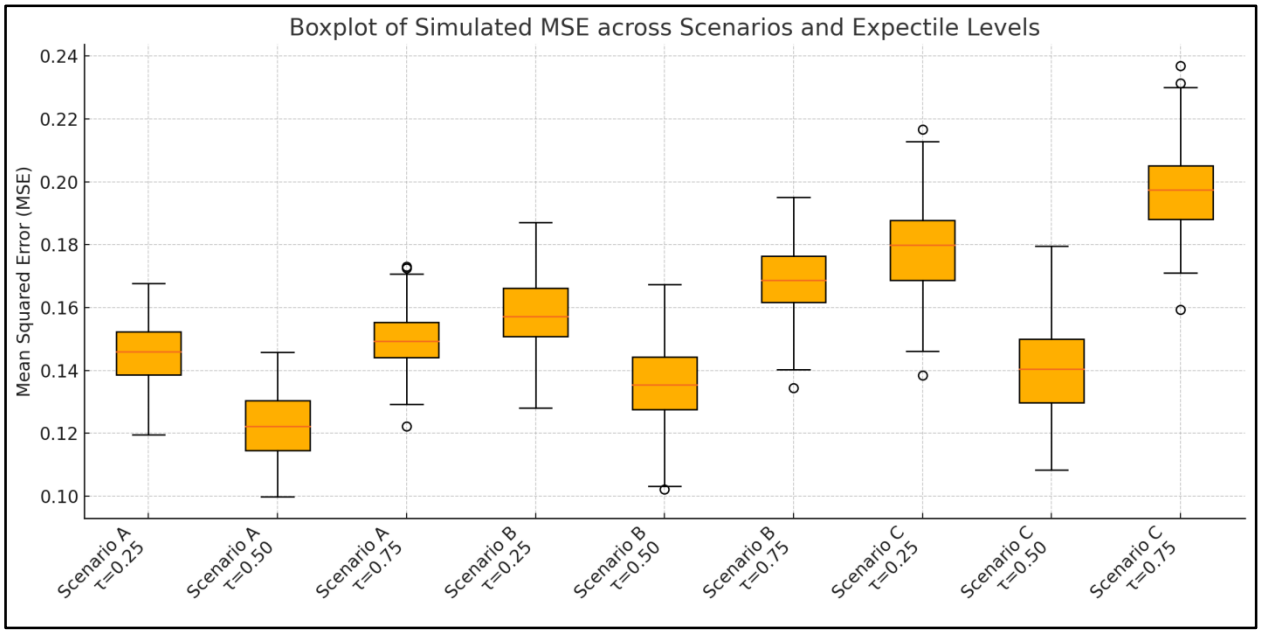

Figure 1: Boxplot of MSE Across Scenarios and Expectile Levels

Figure 1 illustrates the distribution of Mean Squared Error (MSE) values across different simulation scenarios and expectile levels (τ = 0.25, 0.50, 0.75). The results show that Scenario A (symmetric and homoscedastic errors) yields the lowest and most stable MSE values, with narrow interquartile ranges across all τ levels. In Scenario B, the presence of heteroscedasticity slightly increases both the median MSE and variability. Scenario C, which involves skewed error terms, exhibits the highest MSE values and the widest spread, especially at $\tau = 0.75$, indicating that model performance deteriorates under skewed conditions, particularly in the upper tail of the distribution.

4. Real Data Analysis

This section presents the empirical application of Expectile Regression to examine the influence of selected economic, political, and social variables on the Sustainable Development Index (SDI) in Iraq over the period 2010–2020. The aim is to quantify the extent to which key macro-level indicators contribute to sustainable development, with particular emphasis on the unemployment rate.

The dataset includes annual observations from 2010 to 2020. The dependent variable is a composite Sustainable Development Index (SDI), while the independent variables include unemployment rate, inflation rate, GDP growth rate, government expenditure, political stability index, and education enrollment rate. All variables are standardized before estimation. The data are sourced from the World Bank, Iraq Central Statistical Organization, UNDP, and UNESCO.

Table 3: Description of Variables

Variable	Description	Source
SDI	Sustainable Development Index	Iraq CSO / UNDP
Unemployment Rate (%)	Percent of labor force unemployed	World Bank
Inflation Rate (%)	Annual change in consumer price index	World Bank
GDP Growth Rate (%)	Annual real GDP growth	World Bank
Government Expenditure (%)	Public spending as a percentage of GDP	Iraq Ministry of Finance
Political Stability Index	Ranges from -2.5 (weak) to 2.5 (strong)	World Governance Indicators (WGI)
Education Enrollment Rate (%)	Gross enrollment in secondary and tertiary education	UNESCO

Table 4: Summary Statistics (2010–2020)

Variable	Mean	Std. Dev	Min	Max
SDI	0.602	0.045	0.533	0.676
Unemployment Rate	13.9	1.72	11.5	16.2
Inflation Rate	3.1	2.6	0.2	7.9
GDP Growth Rate	2.7	5.3	-4.3	12.1
Government Expenditure	35.4	3.9	28.8	42.5
Political Stability Index	-1.85	0.22	-2.10	-1.50
Education Enrollment Rate	57.8	4.1	50.2	63.9

Table 4 summarizes the descriptive statistics of the study variables for Iraq from 2010 to 2020.

Here's a brief inte

The Sustainable Development Index (SDI) has a mean of 0.602, with moderate variation, indicating relative stability over the period. The unemployment rate averages 13.9%, with a range between 11.5% and 16.2%, reflecting persistent labor market challenges. Inflation is generally low, averaging 3.1%, but with occasional spikes (up to 7.9%).GDP growth rate shows high variability, ranging from negative (-4.3%) to strong positive growth (12.1%), indicating economic instability. Government expenditure remains within a moderate range (28.8% to 42.5% of GDP), suggesting steady fiscal activity. The political stability index is consistently negative, averaging -1.85, highlighting ongoing governance and security issues. Education enrollment averages 57.8%, with gradual improvement over time, supporting human capital development. This table offers a general view of Iraq's economic and social landscape during the study period.

Expectile Regression is applied to estimate the conditional effects of the independent variables on SDI. Three expectile levels are considered—$\tau = 0.25$, 0.50, and 0.75—to capture behavior in the lower, central, and upper parts of the distribution.

The model is fitted using the expectreg package in R. Standard errors are obtained through nonparametric bootstrap with 1,000 replications due to the small sample size. All variables are standardized to allow direct comparison of coefficients.

Table 5: Expectile Regression Estimates

Variable	$\widehat{\beta}(\tau=0.25)$	$\widehat{\beta}$ ($\tau=0.50$)	$\widehat{\beta}$ ($\tau=0.75$)
Unemployment Rate	**-0.0061**	**-0.0050**	**-0.0032**
Inflation Rate	**-0.0018**	**-0.0009**	**0.0001**
GDP Growth Rate	**0.0045**	**0.0063**	**0.0076**
Government Expenditure	**0.0022**	**0.0028**	**0.0035**
Political Stability Index	**0.0112**	**0.0145**	**0.0189**
Education Enrollment	**0.0031**	**0.0040**	**0.0047**

The unemployment rate has a negative impact on SDI at all expectile levels, with the strongest effect observed at the lower tail ($\tau = 0.25$), suggesting that higher unemployment worsens development conditions in already struggling regions. Political stability has the most significant positive effect, especially at $\tau = 0.75$, implying that its role is more pronounced in sustaining high levels of development. GDP growth contributes positively and consistently across all expectiles, increasing in strength toward the upper tail. Inflation has a small and diminishing effect, nearly neutral at $\tau = 0.75$. Education enrollment and government expenditure show positive associations with SDI at all levels, confirming their structural importance in supporting sustainable development.

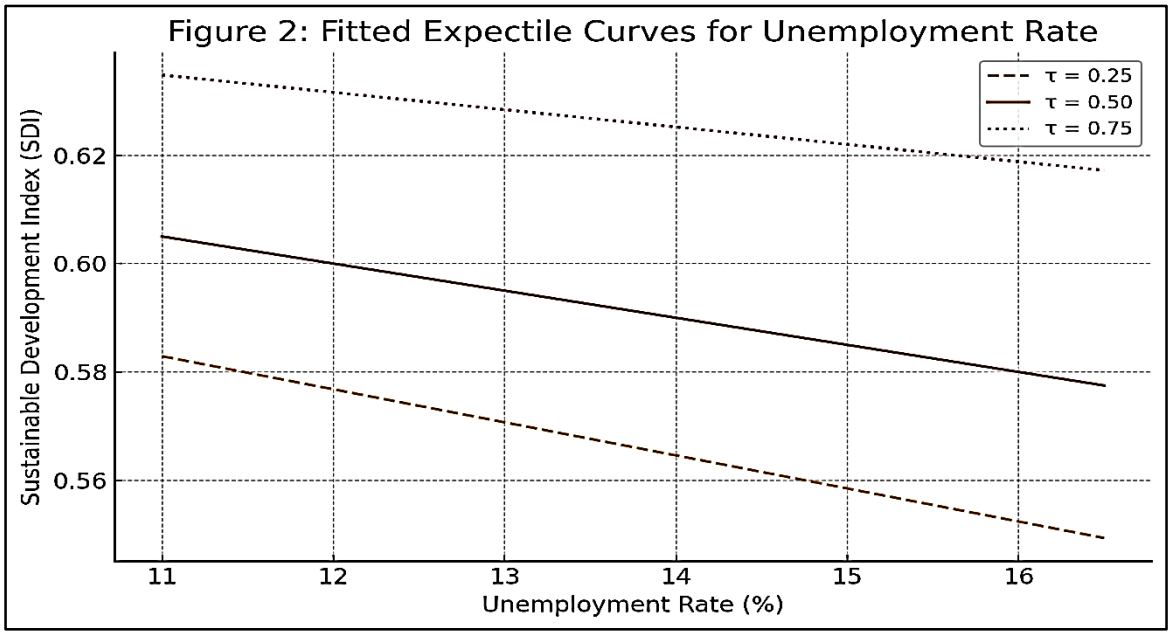

Figure 2: Fitted Expectile Curves for Unemployment Rate

rate and the Sustainable Development Index (SDI) at three levels: $\tau = 0.25$, 0.50, and 0.75. The negative slope across all curves indicates that higher unemployment consistently reduces the level of sustainable development. The steepest decline appears at $\tau = 0.25$, suggesting that the impact of unemployment is strongest in lower-performing development contexts. In contrast, the flatter curve at $\tau = 0.75$ implies a weaker, but still negative, effect at higher development levels.

5. Conclusion

This study investigated the impact of key economic, political, and social factors on sustainable development in Iraq from 2010 to 2020, using Expectile Regression as the primary modeling approach. The findings revealed that unemployment, political stability, government expenditure, and education enrollment significantly influence the Sustainable Development Index (SDI), with varying effects across different levels of the development distribution.
Expectile Regression proved effective in capturing asymmetric relationships, especially the stronger negative impact of unemployment at lower expectiles and the increasing importance of political stability and GDP growth at higher levels. These insights highlight the need for differentiated policy responses tailored to the varying stages of development performance.

The study provides a statistical foundation for policymakers to prioritize strategies that reduce unemployment, enhance education access, and promote institutional stability. While the analysis is based on a relatively small time series, the results demonstrate the practical value of expectile-based modeling in economic development research.

Future work may expand this approach by incorporating regional-level data, additional socio-environmental indicators, or dynamic expectile models to capture time-dependent effects.

References

1. Al-Guraibawi, M., Raheem, S. H., & Mohammed, B. K. (2025). A new modified robust Mahalanobis distance based on MRCD to diagnose high leverage points. *Pakistan Journal of Statistics, 41*(1).
2. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*(366), 427–431.
3. Iraq Central Statistical Organization. (2023). National Development Indicators. Retrieved from http://cosit.gov.iq
4. Ministry of Finance – Iraq. (2023). Annual Budget Reports. Retrieved from https://mof.gov.iq
5. Newey, W. K., & Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica, 55*(4), 819–847.
6. Raheem, S. H., Alhusseini, F. H. H., & Alshaybawee, T. (2020). Modelling volatility in financial time series using ARCH models. *International Journal of Innovation, Creativity and Change, 12*(7), 248–261.
7. Schnabel, S., & Eilers, P. H. C. (2009). Optimal expectile smoothing. *Computational Statistics & Data Analysis, 53*(12), 4168–4177.
8. Sobotka, F., Waltrup, L. S., & Kneib, T. (2021). Expectile regression for economic forecasting. *Applied Economics, 53*(12), 1341–1357.
9. UNDP. (2023). Human Development Reports. Retrieved from https://hdr.undp.org
10. UNESCO. (2023). Education Statistics. Retrieved from https://uis.unesco.org
11. Waltrup, L. S., Sobotka, F., Kneib, T., & Kauermann, G. (2015). Expectile and quantile regression — Similarities and differences. *Statistical Modelling, 15*(6), 499–521.
12. World Bank. (2023a). World Development Indicators. Retrieved from https://data.worldbank.org
13. World Bank. (2023b). Worldwide Governance Indicators. Retrieved from https://info.worldbank.org/governance/wgi/
14. Yu, K., & Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association, 93*(441), 228–237.
15. Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters, 54*(4), 437–447.