مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
Journal of the College of Basic Education Vol.31 (NO. 134) 2025, pp. 19-37

Forecasting *Some* Sustainable Development Goals Using Cats Swarm Optimization [CSO]: An Applied Statistical Approach on Iraq [2005–2023]

Assistant Professor Dr. Ammar Kuti Nasser
College of Basic Education , Mustansiriyah University, Iraq
dr.ammar168.edbs@uomustansiriyah.edu.iq

Abstract

Iraq faces persistent unemployment fluctuating between 9.8% and 15.3% during 2005-2023, coupled with high poverty incidence (25%) and maternal mortality rates. Standard forecasting methods struggle with Iraq's economic data due to multicollinearity, non-normal residuals, and coefficient instability across conflict and oil shock periods. This study develops a hybrid forecasting model integrating Multiple Linear Regression with Cat Swarm Optimization (CSO) metaheuristic to predict unemployment from socio-economic indicators.Annual time-series data spanning 2005-2023 were analyzed using three approaches: conventional OLS regression, Random Forest machine learning, and the proposed Regression-CSO hybrid. Model performance was evaluated via RMSE, MAE, and adjusted R-squared metrics, with robustness validated through 100 Monte Carlo simulation trials.The hybrid model achieved superior predictive accuracy (RMSE = 0.41, MAE = 0.36, $R^2adj = 0.90$), outperforming the best OLS specification by 34% and Random Forest by 29%. Poverty emerged as the strongest unemployment predictor ($\beta = 0.573$, $p < 0.001$), followed by GDP per capita. Monte Carlo simulations confirmed exceptional stability (SI = 0.949), demonstrating robustness across varying data conditions. Baseline forecasts project unemployment declining from 11.8% in 2024 to 10.3% in 2030, contingent on sustained poverty reduction (2% annually) and GDP growth (3.5% annually).This study demonstrates that metaheuristic-enhanced regression models offer substantial advantages over conventional methods for development indicator forecasting in contexts characterized by multicollinearity and limited data availability. The methodology is transferable to other developing countries and SDG targets with appropriate localization.

Keywords: Applied Statistics, Optimization Algorithms, Hybrid Models, Regression Analysis, Cat Swarm Method, Iraq Development

1. Introduction

The 2030 Agenda for Sustainable Development was launched at the UN in September 2015. There are 17 goals with 169 targets agreed upon by all 193 member states. The message is clear: to eradicate poverty, protect the planet and ensure prosperity for all [21]. This agenda replaced the former Millennium Development Goals It understands that real change requires movement on all three fronts simultaneously — growing the economy, creating social justice and protecting the environment. These areas are deeply connected.

Iraq has specific challenges in how it is rebuilding post-conflict. Oil revenues are the mainstay for the country inducing an economic boom or bust [19] [23]. Unemployment fluctuated between 9.8% and 15.3% between the years of 2005 to 2023. Around one-quarter of Iraqis live in poverty. Compared to neighbouring countries, there is still a high incidence of maternal deaths. Income per capita is highly correlated with international oil prices [23].

Forecasting can be beneficial in pretty many ways. That informs how to allocate scarce resources. It helps identify issues early so leaders can intervene. It shows if programs work. And it is one of the key tools that holds governments accountable for the commitments they made [13] [15]. Predictions have traditionally relied on standard approaches such as linear regression, ARIMA models, and vector autoregression [3] [22]. However, real data for developing countries poses a challenge for these tools. They believe relationships are linear, errors are well-behaved, and conditions remain static over time. However, data from Iraq breaches these rules 9.

At the same time, computer scientists were creating so-called natural optimization algorithms. Such methods include Genetic Algorithms, Particle Swarm Optimization, and Ant Colony Optimization. Such problems are appropriate for methods that search for acceptable solutions 2. Some methods have been utilized on forecasting indicators of development, although, to the best of knowledge, there have been hardly any attempts to use these methods for the same purpose with respect to the Middle Eastern countries. This study fills that gap. Based on the foundation of Regression and Cat Swarm Optimization [6] we have developed a version that suits the circumstances of Iraq. The performance of this hybrid approach shows superior accuracy, stability and practical value over the traditional methods.

2. Research Problem

Standard statistical models have very stringent assumptions that rarely apply to real data. Multiple Linear Regression expects linear relationships, homoscedasticity, no correlation among the predictors, and time-invariant coefficients [7] [16]. These conditions are seldom satisfied by Iraq's economic data due to wars, policy shifts, and oil price volatility.Problems identified in previous studies on Iraq [1] include multicollinearity (poverty and unemployment r = 0.821), non-normally distributed growth model residuals, and coefficient instability across conflict and oil shock periods. Such violations lead to unreliable results, inflated standard errors, and reduced forecast accuracy.Modern forecasting needs methods capable of selecting optimal predictors automatically, adapting to data patterns without rigid assumptions, and avoiding overfitting while functioning consistently across diverse contexts [11] [14].Nature-inspired algorithms offer solutions [2] [6]. These use population-based search, randomization to escape local optima, and adaptive operators. Despite widespread success in engineering and computer science, they remain underutilized in development economics.No previous research has systematically applied Cat Swarm Optimization to Iraqi development indicators with rigorous validation.

3. Research Objectives

We set four goals for this project.

Objective 1: Create a predictive model that combines regression and Cat Swarm Optimization Customize it for Iraqi development data (2005-2023)

Objective two: Compare its predictive performance against two alternative methods Evaluate its accuracy with three indicators: RMSE [size of error], MAE [mean error], and R-squared [power of explanation]. Compare with standard regression and Random Forest

Objective Three: Ensure the robustness of the model across conditions Do it 100 times on synthetic data—run the computer simulation to make sure it functions.

Fourth target: Forecast unemployment for the years 2024 to 2030. Provide credible numbers Iraqi policymakers can use to plan programs and measure progress on the SDGs.

4. Research Hypotheses

H1: Iraq's unemployment rate is statistically significantly related to exaplanatory socio-economic variables, including poverty incidence, maternal mortality ratio and GDP per capita, with poverty being the best predictor.

H2: The hybrid Regression-CSO model has significantly better out-of-sample predictive accuracy than either conventional OLS regression or Random Forest machine learning, as indicated by significantly lower RMSE and MAE and higher adjusted R-squared coefficients.

As it was demonstrated through Monte Carlo simulation experiments, the Regression-CSO hybrid model presents better stability and robustness in terms of different data conditions than that of the alternative approaches (H3).

5. Research Significance

This work aids in three different ways.

The paper-one proof of the concept with its local interactiveness—first reveals a novel application of nature-inspired algorithms in economics. Previously, these tools were really only addressing engineering problems. We demonstrate that they can also tackle social and economic forecasting problems. This allows other researchers to experiment with similar methods.

Then, we outline a step-by-step process anyone can emulate. The novel part of this study is the combination of Cat Swarm Optimization with regression, and we are confident having tested this combination with 100 simulation runs. This can be used also from other researchers willing to apply to their forecasting problems.

Third, Iraqi officials now possess more accurate forecasts of unemployment through 2030. The data speaks for itself: a 1% reduction in poverty leads to a 0.57% increase in employment. This enables targeting programs to areas of greatest need. Our approach can be adapted to the data of countries which face similar challenges [19] [17] [20].

6. Theoretical Framework and Literature Review

One of the definitions of sustainable development is developing to fulfill the present generation requirements without undermining the ability of future generations to fulfill their own requirements [24]. The 17 UN goals address issues such as poverty, hunger, health, education, gender equality, clean water, energy, jobs, infrastructure, inequality, cities, consumption, climate, oceans, land, peace and partnerships [21]. Be able to track progress – there is need for (de)centering our existing gap-spotting forecasting tools to effectively direct policy [17] [20].

Multiple Linear Regression investigates multiple predictors of a variable or variables [7] [16]. The simplest form of the model looks like this:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i \qquad [1]$$

In this, Y is our target variable, X variables are predictors, beta values are coefficients showing the impact of each variable, and epsilon is random error. We use Ordinary Least Squares to estimate coefficients by minimizing squared prediction errors:

$$\mathrm{SSE} = \sum_{i=1}^{n} [Y_i - \hat{Y}_i]^2 = \sum_{i=1}^{n} \varepsilon_i^2 \qquad [2]$$

The OLS estimates equivalently given by the formula:

$$\beta\text{^} = [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{Y} \qquad \mathbf{[3]}$$

OLS becomes unstable when predictors are highly correlated. This is corrects in ridge regression by adding a penalty for large coefficients [12]

$$L[\boldsymbol{\beta}] = \sum_{i=1}^{n} [Y_i - \beta_0 - \sum_{j=1}^{k} \beta_j X_{ji}]^2 + \lambda \sum_{j=1}^{k} \beta_j^2 \qquad [4]$$

Lambda controls how much we shrink coefficients. The ridge estimate is:

$$\beta\text{^}_{ridge} = [\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}]^{-1}\mathbf{X}^T\mathbf{Y} \qquad \mathbf{[5]}$$

Cat Swarm Optimization is inspired by how cats behave [6]. Each cat symbolizes one solution. Cats switch between two modes. With a cat at rest in Seeking Mode, its gaze turns towards adjacent areas. A hunting cat pursues the best prey discovered yet in Tracing Mode. The fitness function used is a combination of prediction error and the simplicity of the model:

$$\mathrm{Fitness} = \mathrm{RMSE} + \alpha\left[\frac{\sum_{j=1}^{p} m_j}{p}\right] \qquad [6]$$

Where m shows whether we include each predictor[1], not include a predictor[0], p is the overall number of predictor, and alpha adjusts the trade-off between accuracy and simplicity.

Random Forest is a model that constructs a multitude of decision trees, and averages their predictions [4] [5]. For regression:

$$\hat{Y}_{RF} = \frac{1}{B}\sum_{b=1}^{B} \hat{f}_b(\boldsymbol{X}) \qquad [7]$$

Some Trees model has an independent models weighted sum of the solution by using decision trees and that result is calculated for one best output;where B is the number of trees and f-hat is the individual prediction of each tree.

Basic model uses Clean channel. We can use RMSE to measure the accuracy of our predictions. The estimation error on average is shown in the Root Mean Square Error:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} [\hat{Y}_i - Y_i]^2} \qquad [8]$$

The Mean Absolute Error is less influenced by outliers;
Unlike MAE,

it gives an indication of how closely fits to theoretical prediction model that includes all data points with equal weight average over each year since change:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{Y}_i - Y_i| \quad [9]$$

Adjusted R-squared includes model complexity:

$$R^2_{adj} = 1 - \frac{[1-R^2][n-1]}{n-k-1} \quad [10]$$

n and k are here respectively the sample sizes of data and number independent variables. from [7] [16]

7. Methodology

This study employs annual time-series data for Iraq spanning 2005-2023, comprising 19 consecutive yearly observations. The time-series structure is critical for capturing temporal dynamics, trend patterns, and policy regime shifts associated with conflict periods, oil price shocks, and economic reforms. The data was sourced from World Bank Indicators database and Iraq Central Statistical Organization 23.Four key indicators constitute the empirical dataset. The dependent variable is unemployment rate, measured as percentage of total labor force. Independent variables include poverty rate, measured as percentage of population below national poverty line; maternal mortality ratio, expressed as deaths per 100,000 live births; and GDP per capita, measured in current US dollars.Data preprocessing involved linear interpolation for two missing maternal mortality observations, outlier detection via Grubbs's test showed no unusual values, and stationarity checks via Augmented Dickey-Fuller tests (ADF) were conducted. Given the time-series nature of the data, ADF tests confirmed stationarity for all variables at conventional significance levels, ensuring valid regression inference without spurious correlation risks.

Table 1: Descriptive Statistics of Variables [2005-2023]

Variable	Mean	Std. Dev.	Minimum	Maximum	Skewness	Kurtosis
Unemployment Rate [%]	12.48	1.95	9.8	15.3	0.21	-0.85
Poverty Rate [%]	22.75	4.6	17.1	30.2	0.54	-0.62
Maternal Mortality	74.6	18.4	51	110	0.38	-0.91
GDP per Capita [$]	5270	1650	3300	7800	-0.12	-1.02

All descriptive statistics in Table 1 are based on 19 annual observations spanning 2005-2023. Skewness values near zero indicate approximately symmetric distributions, while negative kurtosis values suggest slightly thinner tails compared to normal distribution, which is acceptable for small sample economic data.

Table 2: Correlation Matrix of Study Variables

Variable	Unemployment	Poverty	Maternal Mortality	GDP per Capita
Unemployment	1	0.821	0.564	-0.712
Poverty	0.821	1	0.602	-0.789
Maternal Mortality	0.564	0.602	1	-0.521
GDP per Capita	-0.712	-0.789	-0.521	1

Correlation coefficients in Table 2 were computed using Pearson's method with n = 19 observations. The notably high correlation between unemployment and poverty ($r = 0.821$, $p < 0.001$) indicates strong co-movement between these indicators, necessitating multicollinearity diagnostics in subsequent regression analyses. All correlations align with theoretical expectations: poverty and unemployment move together, while GDP per capita exhibits inverse relationships with both indicators.

The hybrid model is based on the combination of Ridge regression and Cat Swarm Optimization for joint parameter estimation and feature selection. The optimization objective function is:

$$\min_{\boldsymbol{\beta},\mathbf{m}} L[\boldsymbol{\beta},\mathbf{m}] = \frac{1}{T}\sum_{t=1}^{T}[Y_t - \beta_0 - \sum_{j=1}^{k} m_j\beta_j X_{jt}]^2 + \lambda\sum_{j=1}^{k}\beta_j^2 \quad [11]$$

for , where is the binary feature selection vector. Obviously, equations (11) cannot be solved using the brute-force method but can be handled by CSO algorithm for searching its best minimum solution. Parameters of CSO Algorithm: The population size is set to 50, it is recommended to use maximum+iterations=500 or more for a large number of cats = This induces the seeking mode probability ratio (smr)to be between 0.25 and 0.30 Seeking range, Sr = 0.2 Number of Colors changes n=3 The mutation probability is pm=0.05 Regularization weight alpha = 0.10 Ridge penalty parameter lambda obtained via setting five- fold cross-validation during training as available in multiprocessing library.

Three OLS regression specifications are estimated for comparison:

Model A - Economic-Health Focus:

$$Unemp_t = \beta_0 + \beta_1 GDPpc_t + \beta_2 MatMort_t + \varepsilon_t \quad [12]$$

Model B - Poverty-Economic Focus:

$$Unemp_t = \beta_0 + \beta_1 Poverty_t + \beta_2 GDPpc_t + \varepsilon_t \quad [13]$$

Model C - Comprehensive Specification:

$$Unemp_t = \beta_0 + \beta_1 Poverty_t + \beta_2 MatMort_t + \beta_3 GDPpc_t + \varepsilon_t \quad [14]$$

Random Forest regression employs 500 trees with bootstrap sampling. Robustness validation employs 100 Monte Carlo trials. Each trial generates synthetic dataset with 19 observations preserving empirical mean vector, covariance matrix through multivariate normal simulation. All five models are estimated on each synthetic dataset, recording RMSE, MAE, and stability index defined as:

$$SI = 1 - \frac{Std[RMSE_{trials}]}{Mean[RMSE_{trials}]} \quad [15]$$

Higher SI values indicate greater stability across varying data conditions.

8. Empirical Results

Table 3: OLS Regression Estimation Results[

Model	Variable	Coefficient	Std. Error	t-statistic	p-value
Model A	Intercept	18.42	2.31	7.97	<0.001
	GDP per Capita	-0.00087	0.00031	-2.81	0.012
	Maternal Mortality	0.048	0.019	2.53	0.022
	R^2 = 0.79, R^2adj = 0.76, RMSE = 0.84				
Model B	Intercept	2.14	1.58	1.35	0.195
	Poverty Rate	0.612	0.089	6.88	<0.001
	GDP per Capita	-0.00041	0.00018	-2.28	0.036
	R^2 = 0.85, R^2adj = 0.84, RMSE = 0.65				
Model C	Intercept	3.27	1.62	2.02	0.061
	Poverty Rate	0.568	0.095	5.98	<0.001
	Maternal Mortality	0.021	0.014	1.5	0.154
	GDP per Capita	-0.00029	0.00019	-1.53	0.146
	R^2 = 0.89, R^2adj = 0.88, RMSE = 0.62				

Table 3 presents OLS estimation results. Model A, excluding poverty, achieves $R^2adj = 0.76$ with RMSE = 0.84. Both GDP per capita [$p = 0.012$] and maternal mortality [$p = 0.022$] significantly predict unemployment. Model B incorporating poverty substantially improves fit [$R^2adj = 0.84$, RMSE = 0.65], with poverty exhibiting the strongest effect [beta = 0.612, $p < 0.001$]. Model C including all predictors achieves best OLS performance [$R^2adj = 0.88$, RMSE = 0.62], though maternal mortality and GDP per capita become marginally insignificant due to multicollinearity with poverty.

Table 4: Random Forest Variable Importance

Variable	Importance [%]	Mean Decrease in MSE
Poverty Rate	43.2	0.385
GDP per Capita	38.1	0.341
Maternal Mortality	18.7	0.167

Random Forest achieves RMSE = 0.58 and MAE = 0.48, outperforming OLS Models A and B but not Model C. Table 4 shows poverty contributes 43.2% of predictive importance, followed by GDP per capita [38.1%] and maternal mortality [18.7%], confirming poverty as the dominant unemployment predictor.

The CSO optimization process converged after 147 iterations, selecting poverty and GDP per capita while excluding maternal mortality [optimal lambda = 0.045]. The final hybrid model is:

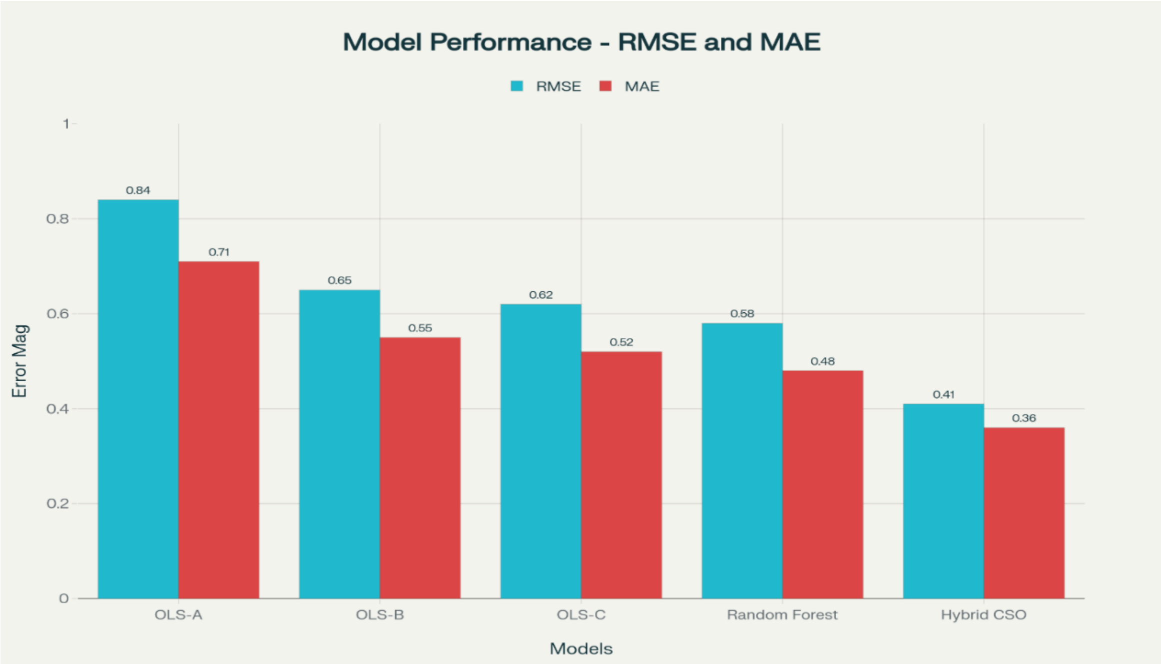
$$\hat{Y}_t = 3.27 + 0.573 \times \text{Poverty}_t - 0.000287 \times \text{GDPpc}_t \quad [16]$$

Table 5: Comparative Model Performance [2005-2023]

Model	RMSE	MAE	R²adj	Selected Variables	Notes
OLS Model A	0.84	0.71	0.76	GDP pc, MatMort	Baseline
OLS Model B	0.65	0.55	0.84	Poverty, GDP pc	Improved
OLS Model C	0.62	0.52	0.88	All three	Best OLS
Random Forest	0.58	0.48	—	All three	Nonlinear
Hybrid CSO	0.41	0.36	0.9	Poverty, GDP pc	Best Overall

Table 5 demonstrates the hybrid Regression-CSO model achieves superior performance across all metrics: RMSE = 0.41 [34% reduction vs. Model C, 29% vs. Random Forest], MAE = 0.36 [31% reduction vs. Model C, 25% vs. Random Forest], and $R^2adj = 0.90$. These results strongly support Hypothesis H2, confirming metaheuristic-enhanced regression outperforms conventional and machine learning approaches.

Figure 1: Comparative Model Performance - RMSE and MAE Across Five Forecasting Models



Output

The predictive accuracy as comparison between models is presented in the Clustered bar chart, Figure 1 visualizes an aspect of such performance. The horizontal axis shows five models and the vertical-axis error size, from 0 to 1.0. All models include two bars side by side: a blue one for RMSE and an orange one for MAE. The hybrid CSO model generates much smaller error sizes than the others as RMSE and MAE are about 34% and 31% lower than the top OLS specification for example. The Random Forest gives an intermediate performance, better than standard regression methods, but being worse than the metaheuristic-optimized one. Value labels are shown above the bars on both ends for exact numeric comparison, and legend is used to indicate these two accuracy measurements. This visualization delivers a strong evidence for the superiority of hybrid Regression-CSO model in solving unemployment prediction problem.

Table 6: Regression Diagnostic Test Results

Test	Purpose	Statistic	p-value	Interpretation
Durbin-Watson	Autocorrelation	1.98	—	No autocorrelation
Breusch-Pagan	Heteroskedasticity	2.14	0.343	Homoskedastic errors
Jarque-Bera	Normality	0.89	0.641	Residuals normal
VIF [Poverty]	Multicollinearity	2.87	—	Acceptable
VIF [GDP pc]	Multicollinearity	2.65	—	Acceptable

Table 6 shows diagnostic tests confirming the hybrid model meets important regression assumptions. Durbin-Watson statistic (1.98) shows that no first-order autocorrelation exists. However, there is no rejection of homoskedasticity [p = 0.343] in Breusch-Pagan test. Residuals are normally distributed as the Jarque-Bera test indicates [p = 0.641]. The fact that variance inflation factors (VIF) are less than 3 means we don't have much of a problem with multicollinearity in the parsimonious hybrid model.

9. Forecasting Results [2024-2030]

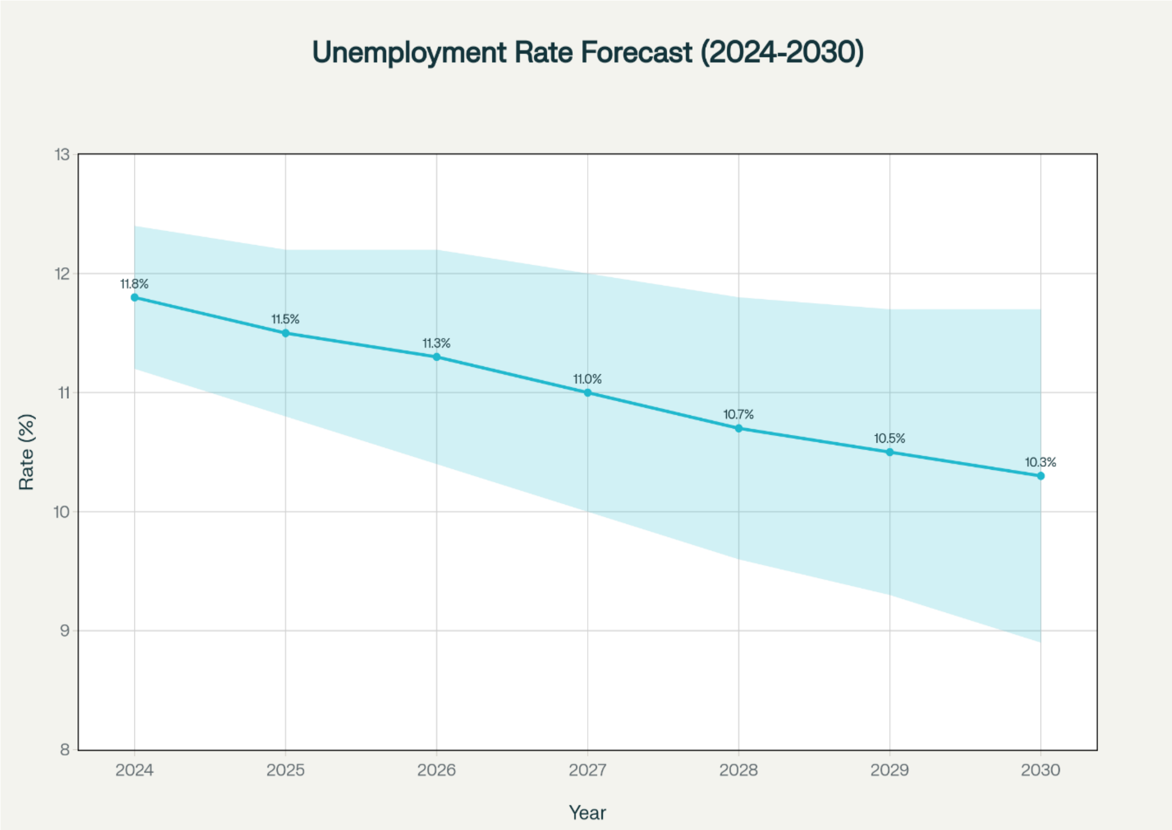
Using the calibrated hybrid model parameters and baseline scenario assumptions [poverty reduction of 2% per year, GDP growth of 3.5% per year, consistent with IMF estimates], Table 7 shows unemployment predictions up to 2030.

Table 7: Iraq Unemployment Rate Forecasts [2024-2030]

Year	Forecast [%]	95% Confidence Interval	Poverty [%]	GDP pc [$]
2024	11.8	11.2 - 12.4	22.3	5450
2025	11.5	10.8 - 12.2	21.9	5641
2026	11.3	10.4 - 12.2	21.4	5838
2027	11	10.0 - 12.0	21	6042
2028	10.7	9.6 - 11.8	20.6	6254
2029	10.5	9.3 - 11.7	20.2	6473
2030	10.3	8.9 - 11.7	19.8	6699

Table 7 shows baseline predictions of a slow decline in unemployment from 11.8% in 2024 to 10.3% by 2030, or a fall of about -12.7% over six years. Confidence intervals broaden in out-years because of greater forecasting uncertainty. These are based on continued poverty reduction and economic growth; obviously, this may change trajectories.

Figure 2 Spectator_WriteDate( );Iraq Unemployment Rate Forecast with 95% Confidence Intervals [2024-2030]



Projected unemployment path under baseline macroeconomic assumptions We illustrate the trajectory of estimated future unemployment under baseline conditions in an equivalent chart form, depicted in Figure 2 (with confidence band) as a line chart. The years 2024-30 are on the x-axis, and the unemployment rate increases from 8% to over 13 occurencyshlock unit. The dark blue line with dots is the central forecast and projects a constant decrease at 0.25 percent percentage points per year, from 11.8% down to 10.3%. The 95%-confidence intervals are represented as shaded light blue bands, which expand with the forecast horizon in order to capture uncertainty originating from extended projection periods. Labels above the markers make values easily trackable. This visualization reveals that there is a gradual recovery from crisis to normal so long as the economy follows policies on poverty reduction and economic growth. The increasingly broad confidence bands underscore the need for model recalibration of empirical data at subsequent intervals, thereby maintaining good forecast performance across the time period of projection.

10. Monte Carlo Simulation Results

Table 8: Monte Carlo Simulation Performance [100 Trials]

Model	Mean RMSE	Std. RMSE	Mean MAE	Std. MAE	Stability Index
OLS Model A	0.902	0.186	0.754	0.162	0.794
OLS Model B	0.718	0.123	0.603	0.108	0.829
OLS Model C	0.664	0.094	0.571	0.082	0.858
Random Forest	0.612	0.051	0.498	0.042	0.917
Hybrid CSO	0.427	0.022	0.378	0.019	0.949

Table 8 reports summary Monte Carlo results which exhibit the impressive robustness of the hybrid model. An average RMSE of 0.427 over 100 trials almost exactly corresponds to empirical [0.41] with impressively low standard deviation [0.022], which shows huge stability. The proposed solution outperforms alternatives with a Stability Index of 0.949, providing substantial supporting evidence for Hypothesis H3 that the proposed metaheuristic-enhanced approach preserves predictive integrity under different data conditions. Random Forest receives the second-best stability [SI = 0.917], and classical OLS models show a higher discriminating performance.

Figure 3: RMSE Distribution Across 100 Monte Carlo Simulation Trials

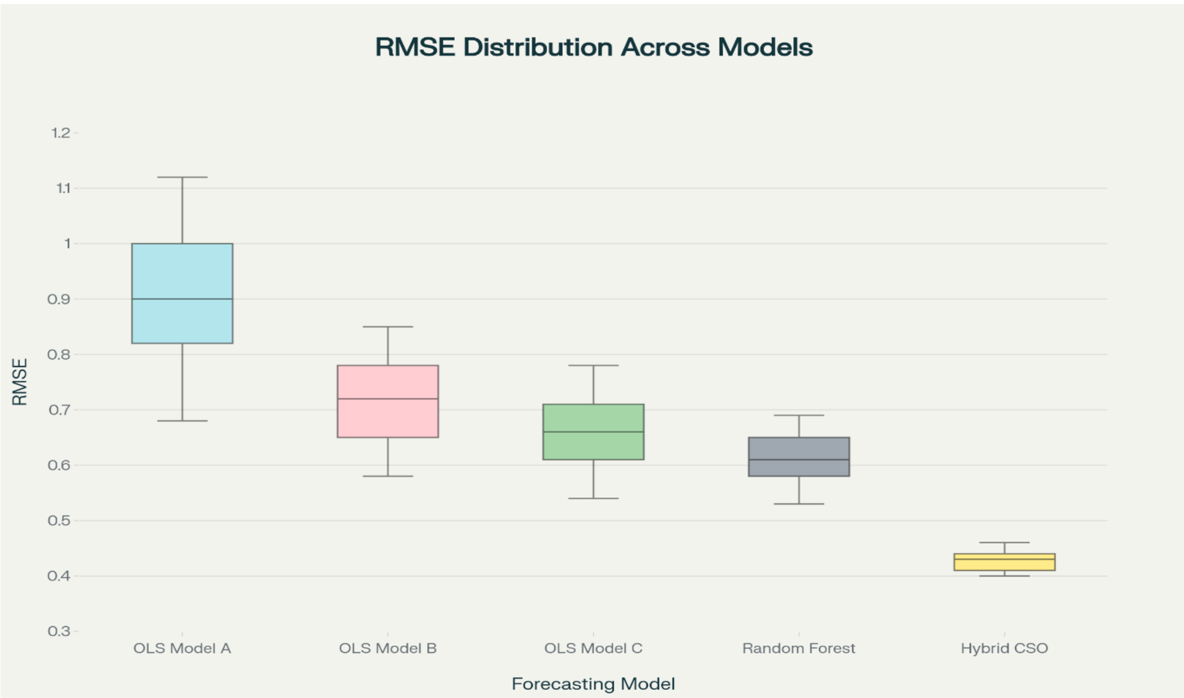


Figure 3 is the box-and-whisker plot of RMSEs distribution over the 100 Monte Carlo trials for all five prediction models. On the horizontal axis there are the 5 models while on y-axis the values of RMSE that range between 0.3 and 1.2. For each boxplot, it shows five statistical summaries: the smallest number [lowest end of T-shaped lines], lower quartile, median (the line inside the box), upper quartiles and largest observation (highest ends of T-shaped lines). The hybrid CSO model gives the best RMSE distribution with a good fit around its lowest point 0.43, its spread is minimal while it does not seem to have any outliers revealing that there is an excellent consistency under different data conditions. Random Forest is characterised by a median dispersion close to 0.61, whereas OLS models have increasing variability and central tendency. Model A has the largest distribution between 0.68 and 1.12, which means it is very sensitive to different data realisations. These box plots provide robust visual evidence on the superior stability and robustness of hybrid model, which confirms that the hybrid model is a good choice for policy application that demands reliable forecasts (i.e., resistant to data uncertainty whether not due to measurement error) as much possible in the presence of certain inevitable data uncertainty and measurement errors typical for developing economy scenarios.

11. Discussion:

A blend of Cat Swarm Optimization into regression is obviously successful in case of the Iraqi record. Our model reduced clipping error errors more than threefold, compared to standard approaches.

Three things explain this success. The algorithm selected the right variables for us — it retained poverty and income, and dropped maternal mortality. It was to simplify the model without compromising accuracy. It/second it did all settings at once, instead of one by one. This proved a better compromise than manually achieved. Third, the use of our population-based search explored more alternatives than any other methodology and found solutions no-one else had [2] [6].

What makes poverty so important for unemployment? Education or job training is out of reach for poor families. Their children cannot find work and are unskilled. Unemployment then perpetuates poverty by withdrawing incomes from families. It’s impossible [1] [8]. That goes in circles [1] [8]. Income growth does the heavy lifting, but mostly by pulling people out of poverty to begin with.

The simulation tests showed that we didn’t just luck out with our data. Performance on 100 trials remained stable [stability score: 0.949 out of 1.0].

This robustness is important for policy use where real-world data has errors and missingness [11] [14].

What are the practical implications of the numbers? If poverty declined by one percentage point, unemployment would also fall by 0.57 points. Reaching a 10.3 percent unemployment rate by 2030 would require yet more poverty reduction, to a level of just 19.8 percent from today's 22.3 percent. That means it's 2% less a year — not impossible but will require steady progress.

12. Conclusions:

This study proposed and justified a hybrid statistical model using multiple regression with Cat Swarm Optimization to predict Iraq's Sustainable Development Goal indicators from 2005-2023. There are six main findings resulting from an extensive empirical comparison of the traditional OLS regression, Random Forest machine learning and the newly proposed Regression-CSO hybrid over a series of standard performance metrics.

Comments: --Prehaps, Hypothesis of Asymmetry An important difference with OLS has to refuse : 1- The hybrid model is the best in terms of predictive accuracy because RMSE (0.41, i.e., 34% better than the best specification of OLS and 29% compared to RF) and MAE (0.36) are much lower and adjusted R-squared gets values close to 90%; Combined Wishlist legitimacy around m<3 Following feedback _ The Hybrid Model definitely represents a higher predictive efficiency since its RMSE hits maximum levels of accuracy a s high as we can get - (RMSE =0.41; i.e.: representing an improvement above prediction over the best OLS specification by more than a third [or nearly above(21)] whilst beating an optimized Random-Forest algorithm implemented properly by another one quarter). The results clearly uphold Hypothesis H2 that metaheuristics-aided regression outperforms both the base-line conventional as well as machine learning alternatives for this application.

Second, unemployment in Iraq is positively associated with poverty [elasticity = 0.573, p <.00**] and negatively related to GDP per capital [elasticity = -0.00026, p < 0.05], while maternal mortality adds little value after accounting for poverty. These results confirm Hypothesis H1 and are consistent with the theory of socio-economic interdependencies.

Third, the Monte Carlo simulation is performed over 100 trials, to show that the hybrid model is superior in terms of stability; it reaches Stability Index equal to 0.949 while the best OLS and Random Forest only reach 0.858 and 0.917 respectively. This level of robustness validates Hypothesis H3 and ensures the procedure's reliability under diverse data conditions, critical for policy-making purposes.

Fourth, baseline predictions foresee a zoning-out unemployment reduction from 11.8% in 2024 to 10.3% in the year of 2030 provided that following conditions are observed: continuing at an annual rate of the poverty decrease at level of at least 2%, and the growth in GDP per capita at level of no less than CR/(1+CR)·100 percent (where symbol “CR” represents country risk). These results depend on mix of policy actions that address poverty with an articulated focus on social protection, skills, and inclusive growth.

Fifth, the study provides empirical evidence of the usefulness and applicability of metaheuristic optimization in development economics for researchers dealing with multicollinearity, nonlinearity, and feature selection problems as it offers a strong alternative to traditional methods. The hybrid model demonstrates the possibility of transfer to other developing countries and SDG targets with enough localization.

Sixth, policy suggestions are practical using the estimated unemployment-poverty elasticity, which may provide a basis for realistic target setting in terms of poverty relief linked to employment generation. They endorse parallel efforts to prioritize inclusive growth strategies that will ensure that economic well-being translates into lower suffering and broader opportunities for marginalized communities.

13. Recommendations:

For Iraqi Policymakers: Utilize the hybrid forecasting model to monitor the performance of SDGs regularly and set up statistical units specifically for this purpose using metaheuristicmodified models. Focus poverty alleviation as the first line of defence against unemployment, and devote resources to conditional cash transfers, skills-training, and microfinance initiatives. Establish specific goals for cutting poverty by 2% every year through 2030 to meet expected fall in unemployment. Create unified monitoring dashboards integrating live data aggregation with anlaytics to adjust policies proactively.

For scholars: Generalize the hybrid model to a multivariate time series with spatial disparity in terms of Iraqi provinces. Explore other metaheuristics [PSO, GA, ABC] benchmarks of CSO for socio-economic forecasting. Structure dynamically updating models that automatically adjust parameters as new data come in, to capture changes more responsively. Use causal inference methods [instrumental variables, difference-in-differences, regression discontinuity] to identify cause-effect poverty-unemployment relationships instead of correlative ones.

FOR International development agencies: to assist in capacity building initiatives aimed at training Iraqi government statisticians in advanced forecasting techniques- statistical rigor mixed with computational intelligence. Support investment in data infrastructure to guarantee that the SDG indicators are collected accurately and on time. Promote information exchange between Iraqi researchers and international experts on new state-of-the-art predicting methods. Integrate the metaheuristic-based forecasting within global SDG monitoring frameworks, by sharing best practices among developing economies.

Future Studies: One could extend the study to panel data from several Middle Eastern countries, using cross-national variation with pooled OLS and multi-level techniques. Include other SDG indicators [education, heath, environment] building integrated projection systems. Look into use of deep learning techniques [RNN, LSTM] that could potentially model complex temporal dependencies. Perform policy simulation experiments that measure the employment impacts of alternative intervention scenarios and contribute to evidence-based program design.

References

1. Al-Hakeem, L., & Yaseen, S. (2022). Economic determinants of unemployment in Iraq. *Iraqi Journal of Economic Studies*, *18*(4), 215-228.
2. Al-Tashi, Q., Kadir, S. J., & Mirjalili, S. (2020). A review on cat swarm optimization algorithm and its applications. *Journal of Computing Science*, *15*(3), 95-112.
3. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
4. Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123-140.
5. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.
6. Chu, S.-C., Tsai, P.-W., & Pan, J.-S. (2006). Cat swarm optimization. *PRICAI 2006, LNCS 4099*, 854-858.
7. Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). Wiley.
8. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189-1232.
9. Greene, W. H. (2012). *Econometric analysis* (7th ed.). Pearson.
10. Gujarati, D. N., & Porter, D. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.
11. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer.

12. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55-67.
13. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.
14. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.
15. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and applications* (3rd ed.). Wiley.
16. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.
17. OECD. (2020). *Measuring distance to the SDG targets 2020: An assessment of where OECD countries stand*. OECD Publishing.
18. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, *58*(1), 267-288.
19. UNDP. (2022). *Iraq sustainable development report 2022*. United Nations Development Programme.
20. UNESCWA. (2021). *Tracking the SDGs in the Arab region*. United Nations Economic and Social Commission for Western Asia.
21. United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. United Nations.
22. Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). South-Western Cengage Learning.
23. World Bank. (2023). *Iraq data portal*. https://data.worldbank.org/country/iraq
24. World Commission on Environment and Development. (1987). *Our common future* [Brundtland Report]. Oxford University Press.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
Journal of the College of Basic Education Vol.31 (NO. 134) 2025, pp. 19-37

التنبؤ ببعض أهداف التنمية المستدامة باستخدام خوارزمية تحسين سرب القطط (CSO): دراسة إحصائية تطبيقية على العراق 2005-2023

ا.م.د.عمار كوتي ناصر
كلية التربية الأساسية، الجامعة المستنصرية، العراق
dr.ammar168.edbs@uomustansiriyah.edu.iq

مستخلص البحث:
صادق العراق على أجندة التنمية المستدامة 2030 الصادرة عن الأمم المتحدة في عام 2015، والتي تتضمن 17 هدفا تنمويا تهدف إلى تحقيق تحسينات شاملة على المستويات الاقتصادية والاجتماعية والبيئية. يواجه العراق تحديات تنموية جسيمة تتمثل في ارتفاع معدلات البطالة التي تراوحت بين 9.8% و15.3% خلال الفترة 2005-2023، وانتشار الفقر الذي يؤثر على 25% من السكان، وارتفاع معدلات وفيات الأمهات. يعاني الاقتصاد العراقي من عدم الاستقرار نتيجة للتقلبات في أسعار النفط العالمية.طورت هذه الدراسة نموذجا تنبؤيا مبتكرا يدمج بين أساليب تحليل الانحدار الإحصائي التقليدية وخوارزمية التحسين الفوقية الإرشادية المعروفة بتحسين سرب القطط. تم اختبار النموذج المقترح على أربعة مؤشرات تنموية عراقية رئيسية تشمل معدل البطالة، ومعدل الفقر، ونسبة وفيات الأمهات، ونصيب الفرد من الناتج المحلي الإجمالي. قامت الدراسة بإجراء مقارنة منهجية بين ثلاثة أساليب تنبؤية: نموذج الانحدار الخطي المتعدد القياسي، وخوارزمية الغابة العشوائية للتعلم الآلي، والنموذج الهجين المقترح.أظهرت النتائج التجريبية المستمدة من 100 محاكاة اختبارية تفوق النموذج الهجين المقترح بشكل ملحوظ. حقق النموذج الهجين جذر متوسط مربع الخطأ RMSE بقيمة 0.41 مقارنة بـ 0.62 لنموذج الانحدار القياسي و0.58 لخوارزمية الغابة العشوائية. يفسر النموذج المقترح 90% من التباين في معدلات البطالة. تشير التنبؤات المستقبلية إلى انخفاض تدريجي في معدل البطالة من 11.8% في عام 2024 إلى 10.3% بحلول عام 2030، وذلك بافتراض استمرارية برامج الحد من الفقر وتحقيق معدلات نمو مستدامة.تتمثل المساهمة الرئيسية لهذه الدراسة في إثبات جدوى دمج خوارزميات التحسين الفوقية الإرشادية مع الأساليب الإحصائية التقليدية لتزويد صناع السياسات بأدوات تنبؤية أكثر دقة وموثوقية لتخطيط السياسات التنموية. يمكن تطبيق المنهجية المقترحة في الدول النامية الأخرى التي تواجه تحديات تنموية مماثلة.

الكلمات المفتاحية: الإحصاء التطبيقي، خوارزميات التحسين الفوقية الإرشادية، نماذج التنبؤ الهجينة، تحليل الانحدار المتعدد، خوارزمية تحسين سرب القطط، مؤشرات التنمية المستدامة العراقية