Journal of Innovative Construction and Petrochemical Technologies, Vol.3, No.2 (Desember, 2026), 506-518

P-ISSN: xxxx- xxxx | E-ISSN: 3007-9861 | DOI: https://doi.org/10.71285/icpt.v3i1

Journal of Innovative Construction and Petrochemical Technologies

https://icpt.ouk.kz/index.php/icpt/index



Research Article

Check for updates

# Development of a Robust Statistical Approach for Selecting Economic Variables Affecting Economic Growth: An Applied Study of the Iraqi Economy

**Omar Makki Rahim**[1*]

Business Administration, College of Administration &Economics, Kirkuk University, Iraq

* Correspondence: omar-macky@uokirkuk.edu.iq

**Abstract:** This study aims to develop a robust statistical approach for selecting economic variables affecting economic growth by integrating the Huber loss function, Adaptive Elastic Net regularization, and stability selection. The proposed approach seeks to improve variable selection accuracy and estimation reliability in the presence of outliers, multicollinearity, and data instability. Its performance was evaluated through Monte Carlo simulation involving 500 replications under four different scenarios, and compared with OLS, LASSO, Elastic Net, Adaptive LASSO, and Huber-AEN methods.

Furthermore, the proposed methodology was applied to the Iraqi economy using annual macroeconomic data obtained from the World Bank for the period 2000–2021. GDP growth was considered the dependent variable, while inflation, foreign direct investment, oil rents, and population growth were included as explanatory variables. The empirical results showed that all four explanatory variables exceeded the predefined stability selection threshold. The simulation findings also indicated that robust estimation techniques reduced coefficient estimation errors under different contamination conditions, although the proposed method did not outperform all competing methods in every scenario.

The study concludes that combining robust estimation, adaptive regularization, and stability selection provides a useful statistical framework for identifying economic variables affecting growth, particularly when dealing with contaminated observations and correlated predictors.

**Keywords:** Robust Regression; Variable Selection; Huber Loss Function; Adaptive Elastic Net; Stability Selection; Monte Carlo Simulation; Economic Growth; Iraqi Economy.

Copyright © Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).



# 1. INTRODUCTION

Economic growth is one of the most important indicators of macroeconomic performance, that is, the capacity of an economy to increase its productive capacity and the standards of living. It is affected by a range of economic variables such as investment, inflation, public spending, foreign direct investment, trade liberalisation, exchange rates, and monetary variables. It is therefore important to know which determinants are more relevant to the understanding of economic performance and to help inform economic policies.

In statistical terms, the problem of feature selection is a difficult one especially if the economic data is multicollinearity, outliers, and correlated observations. The problem can make traditional regression methods like Ordinary Least Squares (OLS) give unstable estimates. Moreover, traditional variable selection methods might include not relevant variables or miss relevant variables which will impact model reliability and predictive performance.

Some recent statistical techniques, such as the Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net and Robust Regression have been helpful with some of these problems. In many cases, however, the individual methods might lead to restricted use in cases of outliers, multicollinearity, and variable selection instability. This inspires the construction of a unified statistical method that is based on stable variable selection, adaptive regularization and strong estimation.

These are important issues to examine empirically because of the characteristics of the Iraqi economy: reliance on oil revenues, economic activity fluctuations and the responses of macroeconomic variables to these fluctuations. A statistical variable selection technique, which is a near advanced method, can be applied to Iraqi economic data to select the most relevant indicators to economic growth and enhance the accuracy of economic statistical estimation.

Based on this, the purpose of this study is to create a strong statistical method in choosing economic variables that impact the economic growth. The proposed approach combines strong loss functions, adaptive Elastic Net regularization and stability selection. It will be tested using Monte Carlo simulation under various data scenarios and contrasted with traditional and state-of-the-art variable selection techniques. Finally, the proposed methodology will be used in economic data for Iraq to determine the most suitable determinants of economic growth.

## 1.1 Research Problem

The main research problem concerns the difficulty of identifying relevant economic variables when explanatory variables exhibit multicollinearity, outliers, and instability. These conditions may reduce the reliability of conventional regression estimates and lead to inconsistent variable selection. Therefore, there is a need for a robust statistical approach capable of improving variable selection accuracy and estimation stability under different data conditions.

## 1.2 Research Objectives

The study aims to:

1. Develop an integrated robust statistical approach for economic variable selection.
2. Combine robust estimation, adaptive Elastic Net regularization, and stability selection.
3. Evaluate the proposed approach using Monte Carlo simulation.
4. Compare its performance with OLS, LASSO, Elastic Net, and Robust Regression.
5. Apply the proposed methodology to identify relevant determinants of economic growth in Iraq.

## 1.3 Research Hypotheses

- H1: The proposed approach improves variable selection accuracy compared with selected conventional methods under specified contamination and multicollinearity conditions.
- H2: The proposed approach provides greater stability in variable selection under repeated sampling.
- H3: The proposed approach reduces estimation and prediction errors under specified simulation scenarios.

## 2. RELATED WORK

### 2.1 Previous Studies

The selection of variables has been an important issue in the statistical modelling as it is important to include relevant explanatory variables in order to make accurate and meaningful models. Traditional regression techniques can fail if the data is multicollinearity, contains outliers or has many explanatory variables. For this reason, a number of penalized and powerful regression techniques have been designed to overcome these difficulties. Kepplinger (2023) suggested an adaptive robust variable selection method, called adaptive PENSE, based on adaptive Elastic Net S-estimators. The proposed approach is a robust estimation and regularization approach to enhance variable selection and estimation of the coefficients when errors are heavy-tailed and contaminated. The results showed that it was effective in maintaining the stability of variables under various contamination levels. Wallisch et al., 2021, examined the stability of variable selection procedures with resampling method. Their study covered the inclusion frequencies as well as the frequencies of model selection. The results highlighted the possible instability in the selection of variables from sample to sample and that the re-sampling methods are valuable tools to assess the reliability of the selected predictors. Kissel and Mentch (2024) presented a forward stability and model path selection technique to explore the uncertainty of choosing one statistical model from a set of plausible models. Their study showed that using the single model chosen was not the most important thing and that they should also take into account alternative models that could give them a comparable predictive performance. Past research in Iraq has focused on various econometric approaches to economic growth. AbduAllahAbed and Saleh (2022) studied the determinants of economic growth in Iraq using the Autoregressive Distributed Lag (ARDL) approach. They looked at multiple macro-economic indicators such as investment, oil revenues, money supply and fiscal indicators. The results emphasized the macroeconomic factors to be important determinants of economic growth. Additionally, Raheem (2025) introduce a powerful variable selection technique using Huber loss and M-estimation with the Reciprocal Elastic Net. It was studied under varying levels of contamination by using simulation and a financial risk application. The findings of this research gives more evidence of the increasing significance of powerful regularization strategies in variable selection.

### 2.2 Research Gap

While a few of these issues have been discussed before – robust regression, penalized variable selection, and model stability – more research is needed to understand their performance in economic applications with correlated predictors and potentially contaminated observations. Specifically, the current research is directed towards designing and testing an integrated statistical approach that integrates robust estimation, adaptive Elastic Net regularization and stability-based variable selection. This study differs from the studies that investigated these techniques individually or applied them to general high dimensional data to focus on evaluation of these techniques under simulated conditions with controlled variables and applying them to macroeconomic variables that influence economic development in Iraq. In this study, the main contribution is to examine if the incorporation of these procedures can enhance the accuracy of variable selection, the reliability of parameter estimates, and the stability of variable selection under various data conditions. The new framework will be tested against other statistical methodologies using Monte Carlo simulation and an empirical test over the Iraqi economy.

## 3. PROPOSED STATISTICAL METHODOLOGY

### 3.1. Model Specification

This study develops a robust variable-selection framework that combines Huber-type robust estimation, adaptive Elastic Net regularization, and a block-based selection-frequency assessment. Robust loss functions reduce the influence of atypical observations (Huber, 1964; Lambert-Lacroix & Zwald, 2011), while Elastic Net and adaptive penalties support sparse estimation and account for groups of correlated predictors (Zou & Hastie, 2005; Zou, 2006).

Consider the following multiple linear regression model:

$$y_t = \beta_0 + \sum_{j=1}^{\rho} X_{tj}\beta_j + \varepsilon_t, \quad t = 1, \dots, n$$

where $y_t$denotes the economic growth rate at time t; $X_{tj}$ is the j-th explanatory variable; $\beta_0$ is the intercept; $\beta_j$ is its regression coefficient; and $\varepsilon_t$ is the error term. In matrix notation:

$$y = \beta_0 1_n + X\beta + \varepsilon$$

Here, y is an n-dimensional response vector, X is an n by p design matrix, beta is the p-dimensional coefficient vector, and $1_n$ is an n-dimensional vector of ones. The explanatory variables are standardized using training-sample statistics, and the intercept is not penalized.

### 3.2. Robust Estimation Using the Huber Loss Function

To limit the effect of large residuals, the Huber loss is used (Huber, 1964; Lambert-Lacroix & Zwald, 2011):

$$\rho_c(u) = \begin{cases} \frac{1}{2}u^2, |u| \le c \\ c|u| - \frac{1}{2}c^2, |u| > c \end{cases}$$

Where c > 0 is a tuning constant. The standardized residual is defined as:

$$u_t = \frac{y_t - \beta_0 - X_t^T\beta}{\hat{\sigma}}$$

where sigma-hat is a robust estimate of residual scale. One possible initial estimate is the scaled Median Absolute Deviation (MAD):

$\hat{\sigma} = 1.4826\, median_{t=1,\dots,n}(|r_t - median(r)|)$

where $r_t$ denotes residuals from an initial robust fit. The factor 1.4826 gives consistency with the standard deviation under Gaussian errors. In the proposed implementation, the scale estimate is obtained from the initial robust fit and held fixed during each penalized optimization step.

### 3.3. Adaptive Elastic Net Regularization

The Elastic Net combines L1 and L2 penalties and can encourage a grouping effect among correlated predictors (Zou & Hastie, 2005). Adaptive weighting assigns different penalty strengths to different coefficients, following the adaptive regularization principle of Zou (2006). Define:

$$w_j = \frac{1}{(|\tilde{\beta}_j| + \delta)^{\gamma}} \quad , \qquad j = 1, \dots, p$$

Where $\tilde{\beta}_j$ an initial robust coefficient estimate, delta > 0 is prevents division by zero, and gamma > 0 controls the degree of adaptivity. The proposed Huber-Adaptive Elastic Net estimator is obtained by minimizing:

$$minimize_{\beta_0,\beta}\left\{\frac{1}{2}\sum_{t=1}^{n} p_c\left(\frac{y_t - \beta_0 - X_t^T\beta}{\hat{\sigma}}\right) + \lambda\alpha\sum_{j=1}^{\rho} w_j|\beta_j| + \frac{\lambda(1-\alpha)}{2}\sum_{j=1}^{\rho}{\beta_j}^2\right\}$$

Where lambda >= 0 controls the overall penalty strength and alpha in (0,1] determines the relative contributions of the L1 and L2 penalties. The L1 component promotes sparsity, while the L2 component helps stabilize estimates when predictors are correlated. Robust adaptive regularized regression has also been studied by Lambert-Lacroix and Zwald (2011) and Kepplinger (2023); therefore, the present study does not claim that robust adaptive regularization alone is new. Its intended extension is the assessment of selection stability using time-ordered blocks in an economic time-series setting.

### 3.4. Block-Based Stability Assessment

Variations in variable selection may exist among samples. Meinshausen and Buhlmann (2010) proposed the stability selection, which is based on repeated subsampling, to evaluate variable selection stability. The resampling units in this study are not individual observations, but rather consecutive blocks since this application involves time-ordered observations. The motivation

behind block resampling is methods for dependent stationary observations (Kunsch, 1989). The choice selector for the variable j in repetition b is the following:

$$I_j^{(b)} = \begin{cases} 1, & \left|\hat{\beta}_j^{(b)}\right| > \varepsilon \\ 0, & |\hat{\beta}_j^{(b)}| \le \varepsilon \end{cases}$$

Where epsilon is a small numerical tolerance. The selection frequency is:

$$\hat{\pi}_j = \frac{1}{B}\sum_{b=1}^{B} I_j^{(b)}$$

The final selected set is defined as:

$$\hat{S}_\pi = \{j : \hat{\pi}_j \ge \pi_{thr}\}$$

Prespecified is the threshold $\hat{\pi}_j$ (in this case 0.80) and then looked at by sensitivity analysis. The block length and subsample design should also be reported. This is not a classical false-selection error guarantee that is given for this adaptation of the time-series block.

### 3.5. Tuning Parameter Selection

These parameters, lambda, alpha and gamma, are chosen by rolling-origin validation. The model is only fitted to the observations before the validation period at each validation origin. This is treating the temporal order and not using future observations to estimate the past models (as accomplished by Bergmeir et al., 2018; Hyndman & Athanasopoulos, 2021). For the selected parameter combination the minimum is achieved of:

$$(\hat{\lambda}, \hat{\alpha}, \hat{\gamma}) = argmin_{\lambda,\alpha,\gamma} \frac{1}{k}\sum_{k=1}^{k} L_k(\lambda, \alpha, \gamma)$$

K is the number of validation folds and $L_k$ is the loss value on the validation fold k. To facilitate comparisons, the validation loss will be the prespecified loss (e.g. mean Huber loss or RMSE), and this will also be done for all competing methods. Within every fold, all pre-processing and adaptive-weight computations are re-computed.

### 3.6. Algorithm of the Proposed Procedure

1. Check for missing observations, outliers, multicollinearity, stationarity and temporal dependence.
2. Identify a regression model based on stationary variables or theoretically appropriate transformations.
3. Normalize predictors with respect to the training set.
4. Get preliminary good coefficient estimates and compute adaptive weights.
5. Estimate the coefficients in Equation (7) by minimizing the expression.
6. Roll origin validate and select tuning parameters.
7. Repeat selection procedure B times, drawing consecutive time blocks.
8. Calculate selection frequencies, using Equation (9).
9. Keep variables as in Equation (10).
10. Retest and test the chosen model using observations not used in the process of developing the model.

### 3.7. Performance Evaluation

Prediction accuracy is assessed using the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE):

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i|$$

Where $m$ is the number of test observations, In simulation experiments, where the true active set is known, variable-selection performance is measured using the True Positive Rate (TPR) and False Positive Rate (FPR):

$$TPR = \frac{TP}{TP + FN}$$

$FPR = \frac{FP}{FP+TN}$

Coefficient estimation accuracy is evaluated using:

$$MAE_{\beta} = \frac{1}{\rho}\sum_{j=1}^{\rho}(\hat{\beta}_j - \beta_j)^2$$

Here, TP, FP, TN, and FN denote true positives, false positives, true negatives, and false negatives, respectively. The coefficient MSE is calculated over all p coefficients, including coefficients that are truly zero.

### 3.8. Methodological Scope

The proposed method fuses the ideas of the established robust loss, of adaptive Elastic Net regularization, and of stability-selection. The specific contribution of this is to explore an approach to selection-frequency adaption based on blocks for economic time-series data and to test the approach under outlier effects, multicollinearity and temporal dependence. The contribution will be evaluated empirically, against the known alternatives, and not presumed to be superior or to have no theoretical error-control guarantees.

## 4. MONTE CARLO SIMULATION

The performance of variable-selection, coefficient-estimation and forecasting is assessed using a Monte Carlo experiment under four representative data conditions. 500 replications of each condition were performed. The model has correlated predictors, serial correlation in the error term, and the error term is contaminated with error in the response and/or explanatory variables. The experiment is designed as a controlled statistical evaluation and the observations generated in the experiment are not real Iraqi economic data.

### 4.1. Simulation design

The sample size was n = 200, with p = 12 candidate predictors. Four predictors were truly relevant, with coefficients (1.5, 1.0, 0.8, 0.5), while the remaining eight coefficients were zero. Predictors were generated from a Gaussian autoregressive correlation structure across columns. The disturbance followed an AR(1) process with parameter φ. The first 160 observations were used for model fitting and the final 40 observations were held out for forecasting evaluation.

$$\varepsilon_t = \emptyset\varepsilon_{t-1} + u_t\,, u_t \sim N(0, \sigma^2)$$

$$y_t = \beta_0 + \sum_{j=1}^{\rho}\beta_j X_{tj} + \varepsilon_t$$

The four implemented scenarios were selected from the proposed parameter ranges: (S1) clean data, ρ = 0.20 and φ = 0.00; (S2) 5% response contamination, ρ = 0.70 and φ = 0.40; (S3) 5% predictor contamination, ρ = 0.70 and φ = 0.40; and (S4) 10% combined predictor and response contamination, ρ = 0.90 and φ = 0.70. Contaminated values were perturbed by eight standard deviations. Each scenario used R = 500 replications.

### 4.2. Methods and implementation

The comparison included OLS, LASSO, Elastic Net, Adaptive LASSO, Huber-based Adaptive Elastic Net (Huber-AEN), and the proposed RAEN-SS procedure. For RAEN-SS, moving blocks of length 8 were resampled 20 times within each replication. A predictor was retained when its selection frequency was at least 0.60. The robust adaptive fitting step was implemented using iterative Huber reweighting followed by adaptive elastic-net shrinkage. Accordingly, the reported results evaluate this stated IRLS-based computational implementation; they should not be described as the output of a different exact numerical optimizer.

$$\hat{\pi}_j = \frac{1}{B}\sum_{b=1}^{B} l\{j \in \hat{S}^{(b)}\}$$

Performance was measured using TPR, FPR, F1-score, coefficient MSE, RMSE, MAE, and the average number of selected predictors. The reported values are averages over the 500 replications. Monte Carlo standard errors (MCSE) are provided for TPR and RMSE in the

accompanying spreadsheet.

$$MAE = \frac{1}{n_{test}} \sum_{t \in T_{test}}^{m} |y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{n_{test}} \sum_{t \in T_{test}} (y_i - \hat{y}_i)^2}$$

$$MAE_\beta = \frac{1}{\rho} \sum_{j=1}^{\rho} (\hat{\beta}_j - \beta_j)^2$$

$$TPR = \frac{TP}{TP + FN} \quad , FPR = \frac{FP}{FP + TN}$$

### 4.3. Variable-selection results

The results of the variable selection are summarized in Table 4. In the clean scenario, LASSO achieved the highest F1-score (0.939) and all methods recovered all four of the relevant predictors on average. The results for response contamination showed Huber-AEN to have the best F1-score (0.685), with RAEN-SS (0.638) following closely, suggesting that the robust fitting method produced superior results for the relevant variables recovered and false variables selected. Predictor contamination results in a decrease of coefficient errors for both Huber-AEN and RAEN-SS, with the latter not being the lowest FPR. Huber-AEN performed better in combined severe contamination (F1-score 0.574) compared to RAEN-SS (F1-score 0.536). In these experiments, therefore, the proposed stability stage was not the predominant one in all the benchmarks.

| Scenario | Method | TPR | FPR | F1 | Selected p |
|---|---|---|---|---|---|
| Clean | OLS | 1.000 | 0.239 | 0.819 | 5.91 |
| Clean | LASSO | 1.000 | 0.072 | 0.939 | 4.58 |
| Clean | Elastic Net | 1.000 | 0.166 | 0.868 | 5.33 |
| Clean | Adaptive LASSO | 1.000 | 0.128 | 0.893 | 5.02 |
| Clean | Huber-AEN | 1.000 | 0.247 | 0.811 | 5.98 |
| Clean | RAEN-SS | 1.000 | 0.194 | 0.847 | 5.55 |
| Response contamination 5% | OLS | 0.963 | 0.916 | 0.508 | 11.18 |
| Response contamination 5% | LASSO | 0.910 | 0.771 | 0.529 | 9.81 |
| Response contamination 5% | Elastic Net | 0.941 | 0.844 | 0.520 | 10.52 |
| Response contamination 5% | Adaptive LASSO | 0.852 | 0.703 | 0.524 | 9.03 |
| Response contamination 5% | Huber-AEN | 1.000 | 0.484 | 0.685 | 7.88 |
| Response contamination 5% | RAEN-SS | 1.000 | 0.592 | 0.638 | 8.73 |
| Predictor contamination 5% | OLS | 0.994 | 0.526 | 0.663 | 8.18 |
| Predictor contamination 5% | LASSO | 0.994 | 0.194 | 0.847 | 5.53 |
| Predictor contamination 5% | Elastic Net | 0.997 | 0.332 | 0.763 | 6.64 |
| Predictor contamination 5% | Adaptive LASSO | 0.990 | 0.234 | 0.814 | 5.83 |
| Predictor contamination 5% | Huber-AEN | 0.994 | 0.391 | 0.728 | 7.10 |
| Predictor | RAEN-SS | 0.992 | 0.401 | 0.720 | 7.17 |

| | | | | | |
|---|---|---|---|---|---|
| contamination 5% | | | | | |
| Combined contamination 10% | OLS | 0.948 | 0.921 | 0.500 | 11.16 |
| Combined contamination 10% | LASSO | 0.894 | 0.814 | 0.508 | 10.09 |
| Combined contamination 10% | Elastic Net | 0.933 | 0.880 | 0.505 | 10.77 |
| Combined contamination 10% | Adaptive LASSO | 0.838 | 0.728 | 0.508 | 9.17 |
| Combined contamination 10% | Huber-AEN | 0.899 | 0.629 | 0.574 | 8.63 |
| Combined contamination 10% | RAEN-SS | 0.930 | 0.778 | 0.536 | 9.95 |

### 4.4. Estimation and forecasting results

The clean-data results show relatively low forecasting errors across all methods, with RAEN-SS yielding RMSE = 0.821 and MAE = 0.659. Under response contamination, robust procedures reduced coefficient MSE markedly compared with OLS and the non-robust penalized methods; Huber-AEN and RAEN-SS produced coefficient MSE values of 0.015 and 0.014, respectively. Under predictor contamination, RAEN-SS attained the lowest coefficient MSE (0.019), RMSE (1.289), and MAE (0.849) among the compared methods. Under combined contamination, Huber-AEN and RAEN-SS again had much lower coefficient MSE than the non-robust methods, but their forecasting errors remained high because the held-out segment also contained contaminated observations.

| Scenario | Method | Coef. MSE | RMSE | MAE |
|---|---|---|---|---|
| Clean | OLS | 0.005 | 0.839 | 0.674 |
| Clean | LASSO | 0.003 | 0.825 | 0.663 |
| Clean | Elastic Net | 0.004 | 0.830 | 0.667 |
| Clean | Adaptive LASSO | 0.003 | 0.824 | 0.662 |
| Clean | Huber-AEN | 0.004 | 0.832 | 0.669 |
| Clean | RAEN-SS | 0.003 | 0.821 | 0.659 |
| Response contamination 5% | OLS | 0.661 | 5.905 | 2.746 |
| Response contamination 5% | LASSO | 0.441 | 5.835 | 2.604 |
| Response contamination 5% | Elastic Net | 0.430 | 5.831 | 2.599 |
| Response contamination 5% | Adaptive LASSO | 0.533 | 5.865 | 2.663 |
| Response contamination 5% | Huber-AEN | 0.015 | 5.582 | 2.018 |
| Response contamination 5% | RAEN-SS | 0.014 | 5.582 | 2.014 |
| Predictor contamination 5% | OLS | 0.056 | 1.393 | 0.937 |
| Predictor contamination 5% | LASSO | 0.047 | 1.369 | 0.922 |
| Predictor contamination 5% | Elastic Net | 0.044 | 1.366 | 0.923 |
| Predictor contamination 5% | Adaptive LASSO | 0.049 | 1.376 | 0.921 |
| Predictor contamination 5% | Huber-AEN | 0.021 | 1.297 | 0.854 |
| Predictor contamination 5% | RAEN-SS | 0.019 | 1.289 | 0.849 |

| | | | | |
|---|---|---|---|---|
| Combined contamination 10% | OLS | 1.905 | 10.108 | 5.119 |
| Combined contamination 10% | LASSO | 1.490 | 9.955 | 4.953 |
| Combined contamination 10% | Elastic Net | 1.306 | 9.885 | 4.901 |
| Combined contamination 10% | Adaptive LASSO | 1.735 | 10.047 | 5.038 |
| Combined contamination 10% | Huber-AEN | 0.306 | 9.447 | 4.136 |
| Combined contamination 10% | RAEN-SS | 0.307 | 9.448 | 4.137 |

### 4.5. Discussion

The simulation provides qualified support for robust estimation when response or predictor contamination is present: Huber-based procedures generally reduced coefficient-estimation error relative to conventional OLS and penalized least-squares methods. The RAEN-SS stability-selection stage was not consistently superior in terms of F1-score or false-positive control in Huber-AEN. It showed mixed performance in the cases of severe combined contaminations. Thus the results cannot be interpreted to suggest universal superiority. Rather, they recommend that the strong coupling of the fitting component is useful and the stability threshold, block length and selection rule need to be further calibrated. Results should be viewed as evidence from four representative scenarios, not as a full-factorial evaluation of all the combinations in the original parameter grid.

## 5. EMPIRICAL APPLICATION: THE IRAQI ECONOMY

### 5.1. Data, variables, and empirical strategy

The empirical application deals with economic factors related to real GDP growth of Iraq. The desired study range is 2000-2023. However, the explanatory indicators selected are not fully comparable as they are only covered on an annual basis until 2021, especially for oil rents. The balanced estimation sample thus is limited to the 2000–2021 period (i.e. 22 level observations), while the regression sample is limited to 21 observations (2001–2021) after first-differencing the explanatory variables. This limitation prevents the creation of or the assumption of observations that are not available.

The dependent variable is GDP growth (annual %). Four candidate explanatory variables are considered: consumer-price inflation (annual %), foreign direct investment net inflows (% of GDP), oil rents (% of GDP), and population growth (annual %). The series are taken from World Development Indicators (WDI), World Bank. The GDP-growth definition is the annual percentage growth rate of GDP at market prices based on constant local-currency values (World Bank, n.d.).

The Augmented Dickey–Fuller test, using a deliberately limited lag order because of the short sample, indicates that GDP growth is stationary in levels, while the explanatory series are not stationary in levels at the 5% level. Accordingly, the explanatory variables enter the empirical regression in first differences. The estimated relationship is:

$$GDPG_t = \beta_0 + \beta_1 \Delta lNF_t + \beta_2 \Delta FDl_t + \beta_3 \Delta OlL_t + \beta_4 \Delta POP_t + \varepsilon_t$$

Where Δ denotes the annual change in an explanatory indicator. The proposed robust adaptive Elastic Net is estimated using Huber loss and adaptive penalties; RAEN-SS additionally applies moving-block stability selection. Because the annual sample is small, the empirical results are interpreted as exploratory evidence rather than definitive causal effects.

### 5.2. Variable definitions

| Symbol | Indicator | Unit | Expected channel |
|---|---|---|---|
| Y | GDP growth | Annual % | Dependent variable |
| INF | Inflation | Annual % | Price stability / uncertainty |
| FDI | FDI net inflows | % of GDP | External investment |
| OIL | Oil rents | % of GDP | Oil dependence / |

| | | | |
|---|---|---|---|
| | | | resource rents |
| POP | Population growth | Annual % | Labour supply / demand pressure |

Source: World Bank, World Development Indicators. Indicator codes: NY.GDP.MKTP.KD.ZG; FP.CPI.TOTL.ZG; BX.KLT.DINV.WD.GD.ZS; NY.GDP.PETR.RT.ZS; SP.POP.GROW

### 5.3. ANNUAL DATA USED IN THE EMPIRICAL APPLICATION

Table 5.1 reports the matched annual observations for 2000–2021. The values are reproduced from World Bank-sourced indicator series. The GDP-growth series and the explanatory indicators are not seasonally adjusted because the data are annual.

| Year | GDP growth | Inflation | FDI (% GDP) | Oil rents (% GDP) | Population growth |
|---|---|---|---|---|---|
| 2000 | 16.92 | 4.98 | -0.000 | 48.58 | 3.34 |
| 2001 | 1.76 | 16.37 | -0.018 | 49.20 | 3.18 |
| 2002 | -8.20 | 19.32 | -0.001 | 46.67 | 3.21 |
| 2003 | -36.66 | 33.62 | 4.562 | 53.00 | 3.05 |
| 2004 | 53.38 | 26.96 | 0.819 | 64.66 | 2.88 |
| 2005 | 1.67 | 36.96 | 1.032 | 65.16 | 2.97 |
| 2006 | 5.65 | 53.23 | 0.588 | 63.56 | 0.72 |
| 2007 | 1.89 | -10.07 | 1.094 | 52.93 | -0.85 |
| 2008 | 8.23 | 12.66 | 1.410 | 55.38 | 1.93 |
| 2009 | 3.38 | 6.87 | 1.431 | 39.11 | 3.60 |
| 2010 | 6.40 | 2.88 | 1.008 | 43.26 | 3.17 |
| 2011 | 7.55 | 5.80 | 1.121 | 50.93 | 3.50 |
| 2012 | 13.94 | 6.09 | 1.560 | 49.61 | 4.49 |
| 2013 | 7.63 | 1.88 | -0.995 | 45.33 | 4.67 |
| 2014 | 0.20 | 2.24 | -4.455 | 46.60 | 3.50 |
| 2015 | 4.72 | 1.39 | -4.542 | 37.26 | 2.71 |
| 2016 | 13.79 | 0.56 | -3.755 | 32.97 | 2.46 |
| 2017 | -1.82 | 0.18 | -2.688 | 38.92 | 2.36 |
| 2018 | 2.63 | 0.37 | -2.149 | 45.77 | 2.42 |
| 2019 | 5.51 | -0.20 | -1.316 | 40.23 | 2.37 |
| 2020 | -12.04 | 0.57 | -1.580 | 27.04 | 2.36 |
| 2021 | 1.58 | 6.04 | -1.270 | 42.79 | 2.27 |

Note: FDI and oil-rent values are expressed as percentages of GDP. A negative FDI value indicates net disinvestment/outflows. Common-coverage data are used for the values shown, and the full set of explanatory data is not available together over the years 2022–2023, so these results are not applied to that period of the multivariate regression.

Each series in the model is transformed:

$$\Delta X_t = X_t - X_{t-1}$$

So, in this case, GDP growth in year t is the dependent variable, and the explanatory variables are the respective changes of inflation, FDI intensity, oil rents and population growth, all measured on an annual basis. The resulting estimation sample is 2001–2021 (n = 21).

### 5.4. DESCRIPTIVE STATISTICS AND PRE-ESTIMATION TESTS

Table 5.2 summarizes the variables in levels over 2000–2021. GDP growth exhibits substantial dispersion, reflecting major economic disruptions and recovery episodes. Inflation also displays marked variation, while oil rents remain a large share of GDP, consistent with Iraq's dependence on the oil sector.

| Variable | Mean | Std. dev. | Minimum | Maximum |
|---|---|---|---|---|
| GDP_Growth | 4.460 | 15.401 | -36.658 | 53.382 |
| Inflation | 10.396 | 15.088 | -10.067 | 53.231 |
| FDI_GDP | -0.370 | 2.215 | -4.542 | 4.562 |
| Oil_Rents_GDP | 47.224 | 9.723 | 27.035 | 65.158 |
| Population_Growth | 2.741 | 1.160 | -0.850 | 4.665 |

Table 5.3. ADF unit-root test (constant; fixed low lag order)

| Series | ADF statistic | p-value | Decision (5%) |
|---|---|---|---|
| GDP growth (level) | -6.377 | 0.000 | Stationary |
| Δ Inflation | -4.584 | 0.000 | Stationary |
| Δ FDI/GDP | -3.152 | 0.023 | Stationary |
| Δ Oil rents/GDP | -3.822 | 0.003 | Stationary |
| Δ Population growth | -4.732 | 0.000 | Stationary |

The first-difference transformation is supported for the explanatory indicators by the ADF results. This procedure reduces the risk of estimating a spurious levels relationship, although the short sample limits the power of unit-root tests.

Table 5.4. Multicollinearity diagnostics for differenced predictors

| Predictor | VIF | Assessment |
|---|---|---|
| Change in inflation | 1.325 | No serious VIF concern |
| Change in FDI/GDP | 1.031 | No serious VIF concern |
| Change in oil rents/GDP | 1.161 | No serious VIF concern |
| Change in population growth | 1.163 | No serious VIF concern |

All VIF values are below 2, suggesting that multicollinearity among the differenced explanatory variables is not severe in this estimation sample. The Durbin–Watson statistic from the OLS specification is approximately 2.05, indicating no obvious first-order residual autocorrelation in this small-sample fit.

## 5.5. VARIABLE SELECTION AND ESTIMATION RESULTS

The robust adaptive Elastic Net objective used in the empirical application is:

$$min_{\beta_0,\beta}\left\{\frac{1}{n}\sum_{t=1}^{n}p_c\left(\frac{y_t-\beta_0-X_t^T\beta}{\hat{\sigma}}\right)+\lambda\alpha\sum_{j=1}^{\rho}w_j|\beta_j|+\frac{\lambda(1-\alpha)}{2}\sum_{j=1}^{\rho}{\beta_j}^2\right\}$$

For RAEN-SS, 500 moving-block resamples of length three were generated. The selection threshold was fixed at π-thr = 0.60. The resulting inclusion frequencies were 0.802 for the change in inflation, 0.876 for the change in FDI/GDP, 0.806 for the change in oil rents/GDP, and 0.774 for the change in population growth. All four predictors exceeded the prespecified threshold.

$$\hat{\pi}_j=\frac{1}{B}\sum_{b=1}^{B}l\{j\in\hat{S}^{(b)}\}$$

Table 5.5 compares the fitted models. Coefficients are standardized; RMSE and MAE are calculated in the original GDP-growth percentage-point scale.

| Method | Variables selected | RMSE | MAE | Selection result |
|---|---|---|---|---|
| OLS | 4 | 9.968 | 7.476 | All four |
| LASSO | 4 | 10.158 | 7.290 | All four |
| Elastic Net | 4 | 10.044 | 7.349 | All four |
| Adaptive LASSO | 4 | 10.141 | 7.290 | All four |
| Huber-AEN | 4 | 10.017 | 7.332 | All four |
| RAEN-SS | 4 | 10.017 | 7.332 | All four |

All four of the candidate predictors were retained in the proposed RAEN-SS procedure, as each predictor had a selection frequency greater than 0.60. Its in-sample RMSE was 10.017 and MAE was 7.332. The lowest RMSE value was obtained by the OLS (9.968), and the lowest MAE values (around 7.290) were obtained with LASSO and Adaptive LASSO. Thus, the empirical findings are not evidence of the universal superiority of RAEN-SS. The principal empirical result is that the four candidate variables were consistently selected among the block resamples for the selected threshold.

The standardisation coefficients of the Huber-AEN were −0.181 for inflation, −0.673 for FDI/GDP, 0.214 for oil rents/GDP, and 0.190 for population growth. This is a non-causal association, not a causal effect. The negative coefficient of FDI is to be taken with a pinch of

salt, as the net FDI flows can become negative and the annual sample is of short duration.

One significant restriction is the exceptional GDP-growth movement around 2003-2004, driven by conflict and recovery. While robust estimation decreases the impact of extreme residuals, these extreme residuals could be real structural events and not measurement errors. As such, the results should be seen in conjunction with Iraq's economic history, and not as an argument for its removal.

The general idea of the empirical application is the robust variable selection in macroeconomics of the Iraqi economy. The results must be considered exploratory however, because of the limited common sample (21) regression observations, the short time span and limited indicator coverage. If revised and complete WDI series become available, it is recommended to extend the sample.

## 6. CONCLUSIONS AND RECOMMENDATIONS

### 6.1. Conclusions

1. The study designed and analyzed a powerful statistical model that incorporates Huber-type estimation, adaptive Elastic Net regularization and block-based stability selection in order to uncover economic predictors of GDP growth.
2. Models with robust fitting, in general, lowered the error of the coefficients under the response and predictor contamination in the Monte Carlo experiment. But RAEN-SS was not always as successful as the benchmarks in all scenarios. The results were dependent on type and level of contamination and cannot be claimed as all encompassing.
3. Annual indicators from the World Bank were adopted for the Iraqi application. There was no complete set of explanatory variables from the years 2015 to 2021, so the balanced level data spanned 2000–2021 and first-differencing yielded 21 observations for the final regression. This is a shorter time period than the original 2000-2023.
4. Based on the ADF results, GDP growth in levels and first differences of the explanatory indicators were used. The multicollinearity of the differenced predictors was not a serious problem in the estimation sample as shown by the VIF results.
5. The inflation rate, FDI net inflows to GDP, oil rents to GDP and population growth in all surpassed the prespecified stability threshold of 0.60. The procedure thus did not result in a subset of predictors.Thus, the procedure did not result in a subset of predictors.
6. The proposed method has not been able to get the minimum in sample RMSE in the application of Iraq. The MAE for the robust Huber-AEN/RAEN-SS fit was the smallest, and the RMSE was the smallest for OLS. These findings suggest a compromise between the squared-error fit and resistance to large residuals.
7. The strong GDP growth extreme around 2003-2004 is probably correlated with major historical and structural events. They do not necessarily have to be deemed a "false observation" and it is not acceptable to totally remove the influence of such observations by using robust procedures but to not interpret the economic aspects.

### 6.2. Recommendations

1. Update the empirical data set once consistent World Bank and Iraqi official data exist for 2022-2023 and subsequently re-estimate the model for the entire intended period.
2. The effects of structural breaks and events related to conflict, oil-price shocks, and the COVID-19 era should be studied in future research, and methods appropriate to short annual samples should be used.
3. Keep strong estimation as a complementary analysis and explain its performance in conjunction with OLS, LASSO, Elastic Net and Adaptive LASSO, not as an all-seeing, all-knowing analysis.
4. Conduct a sensitivity analysis for alternative stability thresholds and block lengths, as specified in advance. Provide an indication of the limited precision of selection frequencies with short annual series.

5. Strengthen non-oil productive sectors, enhance the investment climate and mitigate growth vulnerability to oil market fluctuations from an economic standpoint. The directions of these policies are consistent with the empirical framework's significance for oil rents and indicators of investment, but the estimated relationships are not causal.

## REFERENCES


1. Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis, 120, 70–83.
2. Huber, P. J. (1964). Robust estimation of a location parameter. The Annals of Mathematical Statistics, 35(1), 73–101.
3. Kepplinger, D. (2023). Robust variable selection and estimation via adaptive elastic net S-estimators for linear regression. Computational Statistics & Data Analysis, 183, 107730.
4. Kock, A. B., Medeiros, M., & Vasconcelos, G. (2020). Penalized time series regression. In P. Fuleky (Ed.), Macroeconomic forecasting in the era of big data (pp. 193–228). Springer.
5. Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. The Annals of Statistics, 17(3), 1217–1241.
6. Lambert-Lacroix, S., & Zwald, L. (2011). Robust regression through the Huber's criterion and adaptive lasso penalty. Electronic Journal of Statistics, 5, 1015–1053.
7. Meinshausen, N., & Bühlmann, P. (2010). Stability selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(4), 417–473.
8. Wisnowski, J. W., Simpson, J. R., Montgomery, D. C., & Runger, G. C. (2003). Resampling methods for variable selection in robust regression. Computational Statistics & Data Analysis, 43(3), 341–355.
9. World Bank. (n.d.-a). Foreign direct investment, net inflows (% of GDP) – Iraq. World Development Indicators.
10. World Bank. (n.d.-b). GDP growth (annual %) – Iraq. World Development Indicators.
11. World Bank. (n.d.-c). Inflation, consumer prices (annual %) – Iraq. World Development Indicators.
12. World Bank. (n.d.-d). Oil rents (% of GDP) – Iraq. World Development Indicators.
13. World Bank. (n.d.-e). Population growth (annual %) – Iraq. World Development Indicators.
14. Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101(476), 1418–1429.