ECONOMETRICS
EKONOMETRIA
Advances in Applied Data Analysis

Econometrics. Ekonometria.
Advances in Applied Data Analysis
Year 2025, Vol. 29, No. 3
ISSN 2449-9994
journals.ue.wroc.pl/eada

Rethinking Scale in Applied Econometrics: Practical Impacts of Log Transformations on Model Performance

Fatih Chellai

Department of Economics, Ferhat ABBAS University, Sétif, Algeria

e-mail: fatih.chellai@univ-setif.dz

ORCID: 0000-0002-3249-846X

Quote as: Chellai, F. (2025). Rethinking Scale in Applied Econometrics: Practical Impacts of Log Transformations on Model Performance. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, *29*(3), 1-14.

DOI: 10.15611/eada.2025.3.01

JEL: C18, C51, C22

Abstract

Aim: This study examines the practical impact of measurement scale choices, particularly logarithmic transformations, on the accuracy, reliability, and interpretability of econometric models using macroeconomic indicators from the MENA region.

Methodology: Using real-world MENA data, multiple techniques (OLS, Fixed Effects, Random Effects, GMM) were applied with both raw and log-transformed variables to compare outcomes for variance stability, normality, stationarity, and heteroskedasticity.

Results: Log transformations substantially improve model diagnostics by stabilising variance, enhancing normality, and reducing heteroskedasticity, yielding more precise estimates. However, they alter coefficient interpretation, highlighting a trade-off between statistical robustness and economic meaning.

Implications and Recommendations: The findings underscore the critical need for appropriate data transformations to ensure valid and interpretable results, offering practical guidance for researchers and policymakers in model specification for policy analysis and forecasting.

Originality/Value: This study provides a novel comparative exploration of transformation effects across methodologies and one of the few region-specific investigations using MENA data, delivering actionable insights for methodological and applied research.

Keywords: log transformation, econometric modelling, MENA region, stationarity tests, measurement scales, model diagnostics

1. Introduction

Econometric modelling is an essential tool in empirical research across economics, finance, and social sciences, where the goal is to quantify relationships between economic variables, and make inferences based on the observed data. However, the reliability of these inferences is heavily contingent on the proper specification of models, which includes careful consideration of measurement scales for the variables involved (Wooldridge, 2010; Greene, 2018). A pivotal issue in econometric analysis is the choice between using raw or transformed variables, such as logarithmic transformations, which can substantially alter model outcomes and interpretations (Stock, & Watson, 2020).

Logarithmic transformations are often employed to address various econometric concerns, including linearising non-linear relationships, stabilising variance, and normalising the distribution of variables (Wooldridge, 2010). Such transformations are particularly useful when dealing with skewed data or when the relationship between variables is multiplicative rather than additive (Gujarati, & Porter, 2009). For example, in models of economic growth or income distributions, the logarithmic transformation can be instrumental in mitigating heteroskedasticity and enhancing the robustness of econometric estimates (Barro, 1991; Solow, 1956). However, the use of logged variables is not without controversy as the appropriateness of such transformations depends on the underlying data-generating process and the specific context of the analysis (Kennedy, 2008).

A crucial aspect of using logged variables in econometric models is their impact on elasticity interpretation. When variables are logged, the coefficients represent elasticities rather than marginal effects, which provides a more intuitive understanding of percentage changes in the dependent variable with respect to percentage changes in the independent variable (Mankiw, 2014), yet this also introduces complexity, as the interpretation is conditional on the units and scale of measurement, potentially leading to misunderstandings or misinterpretations if not properly accounted for (Angrist, & Pischke, 2009). Whilst logarithmic transformations can mitigate issues related to non-normality and non-constant variance, they may also induce bias if the logarithmic form is not an accurate reflection of the true relationship between variables (Greene, 2018).

The choice of whether to log or not to log variables has further implications for hypothesis testing and inference. For instance, hypothesis tests based on models with logged variables can be sensitive to the choice of transformation, as well as to potential omitted variable bias, measurement errors, and endogeneity issues (Wooldridge, 2010; Cameron, & Trivedi, 2005). These concerns highlight the need for careful consideration and rigorous testing of model specifications, particularly in applied econometric research where the goal is to draw meaningful and policy-relevant conclusions (Stock, & Watson, 2020).

Recent literature has explored the ramifications of using logged versus unlogged variables, revealing mixed findings on their effects on econometric estimates and inferences, e.g. a study by Kennedy (2008) found that while logarithmic transformations can improve model fit and inference in certain cases, they may lead to erroneous conclusions if the log-linear model is misspecified. Conversely, other studies suggested that failing to log-transform highly skewed data can arrive at biased estimates and invalid inference, particularly in cases involving heavy-tailed distributions or outliers (Wooldridge, 2010; Gujarati, & Porter, 2009).

Given these considerations, the impact of measurement scales, including the choice of logged versus unlogged variables, warrants careful scrutiny in econometric modeling. This study aimed to explore these effects systematically, examining how different measurement scales influence econometric estimates and the validity of inferences. By leveraging both theoretical insights and empirical applications, the author contributes to the ongoing discourse on best practices for variable transformation in econometric analysis, providing guidance for researchers on how to navigate the complexities associated with measurement scales and model specification.

2. Literature Review

Measurement scales and their appropriate usage have long been central topics in econometric research, significantly affecting the validity and reliability of model inferences and estimates. The decision to log-transform variables, as opposed to using their raw values, introduces a critical aspect of econometric analysis that impacts the interpretation of coefficients, hypothesis testing, and overall model performance (Wooldridge, 2010). The application of different measurement scales can be traced back to foundational econometric texts that stress the importance of understanding the underlying data generation processes (DGPs) and the nature of relationships between variables (Greene, 2018). Early works in econometrics underlined the necessity of using appropriate transformations to ensure that models are correctly specified and that their results are interpretable (Gujarati, & Porter, 2009).

The theoretical underpinnings of variable transformations, particularly logarithmic transformations, have been widely discussed in the econometric literature. Logarithmic transformations are often recommended for several reasons: they can linearise non-linear relationships, stabilise variance, and make the data more normally distributed (Wooldridge, 2010; Greene, 2018). Theoretical models in economics frequently assume multiplicative relationships between variables, for example the Cobb-Douglas production function is inherently multiplicative, making a log transformation natural and appropriate (Mankiw, 2014). By logging both sides of the equation, researchers can convert a multiplicative model into a linear one, facilitating easier estimation and interpretation (Kennedy, 2008).

The choice of whether to transform variables also depends on the interpretability of the coefficients. In a log-log model, the coefficients represent elasticities which describe the percentage change in the dependent variable resulting from a one percent change in an independent variable (Gujarati,& Porter, 2009). This interpretation is often more intuitive in economic contexts where percentage changes are more meaningful than absolute changes. However, when using raw variables, coefficients indicate marginal effects which may be more appropriate in some contexts, such as policy analysis where absolute changes are of interest (Stock, & Watson, 2020).

Empirical studies provided mixed evidence on the effects of using logged versus unlogged variables. For instance, a study by Barro (1991) on economic growth across countries found that using logged GDP and other economic indicators improved the model's fit and provided more robust inferences compared to using raw variables. This was particularly evident in cases where the data exhibited skewness or where the relationships between the variables were multiplicative in nature. Similarly, Solow (1956) demonstrated that log transformations could better capture the diminishing returns to scale in production functions, thereby aligning more closely with theoretical expectations in economic growth models.

The empirical literature also cautions against the indiscriminate use of log transformations. Studies have shown that in cases where the data are not heavily skewed or where relationships are approximately linear, using raw variables may actually lead to more accurate and unbiased estimates (Angrist, & Pischke, 2009). Furthermore, Kennedy (2008) highlighted that logarithmic transformations could distort the true nature of the relationship between variables if the underlying DGP is not multiplicative. Such distortions could lead to incorrect conclusions, especially in policy-oriented research where precision in inference is crucial (Greene, 2018).

The decision to use logged versus unlogged variables is not without controversy and has sparked considerable debate within the econometric community. One major area of contention is the issue of heteroskedasticity and how it is affected by variable transformations. While log transformations are often used to stabilise variance, they may not always be the optimal choice, particularly in cases where the variance is constant across levels of the independent variable (Wooldridge, 2010) – here a Box-Cox transformation or a different non-linear transformation might be more appropriate (Cameron, & Trivedi, 2005).

Another methodological issue pertains to the interpretation of interaction terms in models with logged variables. When both the dependent and independent variables are logged, the coefficient on an interaction term represents the cross-elasticity of the two variables, which can complicate their interpretation, particularly when trying to disentangle the effects of one variable from another (Wooldridge, 2010). This complexity led some researchers to advocate for models that use a combination of logged and unlogged variables to retain interpretability while benefiting from the advantages of transformation (Stock, & Watson, 2020).

The use of different measurement scales has significant implications for hypothesis testing and inference in econometric models. The choice of transformation affects the distribution of error terms, which in turn influences the validity of standard hypothesis tests, such as t-tests and F-tests (Gujarati, & Porter, 2009). If the log transformation is inappropriate for the data, the resulting model may exhibit residual patterns that violate the assumptions of ordinary least squares (OLS) regression, leading to biased or inconsistent parameter estimates (Greene, 2018).

Recent research also explored the impact of measurement scales on econometric methods that go beyond the OLS framework, for example quantile regression, which estimates the conditional median or other quantiles of the response variable, can be particularly sensitive to the scale of measurement used (Koenker, 2005). It was found that quantile regression models using logged variables can yield different insights into the data’s structure and relationships compared to models using unlogged variables, highlighting the need for careful consideration of scale in advanced econometric techniques (Angrist, & Pischke, 2009).

The impact of measurement scales is also evident in non-linear and panel data models, where the choice between logged and unlogged variables can affect both the estimation and interpretation of model parameters. In models with binary or categorical dependent variables such as logit or probit models, using logged continuous predictors can help linearise the relationship between predictors and the log-odds of the outcome, facilitating easier interpretation and more accurate predictions (Wooldridge, 2010). However, in the context of panel data models where individual-specific effects are accounted for, the use of logged variables can complicate the interpretation of fixed and random effects, particularly when considering within-group versus between-group variation (Greene, 2018).

Recent studies in dynamic panel data modeling also highlighted the importance of scale in determining the properties of estimators, such as the generalised method of moments (GMM) estimators, which rely on instruments to control for endogeneity (Arellano, & Bover, 1995). The choice of transformation can affect the validity of instruments and the consistency of estimators, suggesting that researchers must carefully evaluate the implications of their scale choices in dynamic settings (Blundell, & Bond, 1998).

Given the mixed findings and ongoing debates, a consensus has emerged around the importance of context-specific considerations in choosing whether to log or not to log variables in econometric models. Scholars recommend a data-driven approach that involves testing different specifications and transformations to determine which provides the best fit and most reliable inferences (Stock, & Watson, 2020; Wooldridge, 2010). Diagnostic tests, such as examining residual plots for patterns and conducting heteroskedasticity tests, are essential in guiding these decisions (Gujarati, & Porter, 2009). Moreover, simulation studies have been proposed as a method to explore the potential biases and variances associated with different transformations, providing further insights into the appropriateness of using logged versus unlogged variables in specific contexts (Cameron, & Trivedi, 2005). They underscore the need for transparency in reporting model specifications and the rationale behind the choice of transformations to ensure replicability and credibility in empirical research (Angrist, & Pischke, 2009).

The choice between using logged or unlogged variables in econometric modelling remains a nuanced and complex decision that requires careful consideration of theoretical, empirical, and methodological factors. While logarithmic transformations offer significant benefits in terms of interpretability and model robustness, they are not universally applicable, and their appropriateness must be evaluated on a case-by-case basis (Kennedy, 2008; Wooldridge, 2010). Future research could further investigate the implications of different transformations in a wider range of contexts, including emerging areas such as machine learning and big data analytics, where traditional econometric principles are increasingly intersecting with advanced computational techniques (Stock, & Watson, 2020).

3. Methodology

The methodology section provides a comprehensive overview of the econometric approaches and statistical techniques employed in this study to examine the impacts of measurement scales (logged vs. unlogged variables) on econometric model inferences and estimates. This section covers the theoretical foundations of the econometric models used, the data collection and preparation processes, the specific transformations applied to the variables, and the statistical tests conducted to ensure model validity and robustness.

3.1. Data Collection and Preparation

Data for this study were sourced exclusively from the World Bank's publicly available datasets. The World Bank provides a comprehensive collection of global development data, including a wide range of economic, social, and environmental indicators. For this analysis the author focused on the key economic indicators commonly used in econometric analyses, such as Gross Domestic Product (GDP), inflation rates, interest rates, and unemployment rates. The dataset spanned multiple countries and covered the period from 1991 to 2023, ensuring a comprehensive analysis of the effects of different measurement scales across various economic contexts and time frames.

The collected data underwent a rigorous cleaning process to handle missing values, outliers, and inconsistencies. Missing values were treated using multiple imputation methods to avoid bias in estimates (Little, & Rubin, 2002). Outliers were identified using the interquartile range (IQR) method and were either removed or winsorised, depending on their impact on the overall distribution of the data (Hastie et al., 2009).

3.2. Variable Transformation and Measurement Scales

This study focused on understanding the effects of different measurement scales on econometric inferences. Two primary approaches were taken: models using raw (unlogged) variables and models using logarithmically transformed (logged) variables. The transformation process involved taking the natural logarithm (ln) of the continuous variables of interest, such as GDP, inflation rates, and interest rates, to stabilise variance and normalise the distribution (Wooldridge, 2010).

The choice of natural logarithm over other logarithmic bases (such as log10) is consistent with econometric convention and provides a straightforward interpretation of elasticities in terms of percentage changes (Kennedy, 2008). The formula for transforming variable X using the natural logarithm is given by:

$$Y = \ln(X),$$

where X is the transformed variable and X is the original variable. This transformation is particularly useful when the relationship between variables is multiplicative rather than additive (Gujarati, & Porter, 2009). For instance, when examining economic growth rates, a log transformation can convert a multiplicative model into a linear one, facilitating easier estimation and interpretation.

3.3. Econometric Models and Estimation Techniques

To assess the impact of different measurement scales on econometric model outcomes, several econometric models were specified and estimated using both logged and unlogged variables. The primary models used in this study included:

3.3.1. Ordinary Least Squares (OLS) Regression

The OLS regression model is the foundational econometric technique used to estimate the relationship between a dependent variable and one or more independent variables. For the unlogged model, the general form is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i,$$

Where Y_i is the dependent variable, $X_{1i}, X_{2i}, \ldots, X_{ki}$ are the independent variables, $\beta_0, \beta_1, \ldots, \beta_k$ are the coefficients to be estimated, and ϵ_i is the error term (Wooldridge, 2010).

For the logged model, the equation is:

$$ln(Y_i) = \alpha_0 + \alpha_1\, ln(X_{1i}) + \alpha_2\, ln(X_{2i}) + \cdots + \alpha_k\, ln(X_{ki}) + \nu_i,$$

where $ln(Y_i)$ and $ln(X_{1i}), ln(X_{2i}), \ldots, ln(X_{ki})$ represent the natural logarithms of the dependent and independent variables, respectively, and ν_i is the error term for the log-transformed model (Greene, 2018).

3.3.2. Log-Log Model for Elasticity Estimation

In econometric analyses where the elasticity of relationships is of primary interest, a log-log model is used. Coefficient β_j in a log-log model is interpreted as the elasticity of Y with respect to X_j, which indicates the percentage change in resulting from a one percent change in X_j (Gujarati, & Porter, 2009). The model specification is:

$$ln(Y) = \beta_0 + \beta_1\, ln(X_1) + \beta_2\, ln(X_2) + \cdots + \beta_k\, ln(X_k) + \epsilon.$$

3.3.3. Fixed Effects and Random Effects Models for Panel Data

Given the panel structure of the data (multiple countries over time), both fixed effects and random effects models were employed to account for unobserved heterogeneity across entities (countries) (Baltagi, 2005). The fixed effects model can be specified as:

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it},$$

where Y_{it} is the dependent variable for country i at time t, α_i represents the entity-specific intercept capturing the fixed effects, X_{it} is the independent variable, and ϵ_{it} is the error term (Wooldridge, 2010).

The random effects model, on the other hand, assumes that the entity-specific intercepts are random and uncorrelated with the independent variables:

$$Y_{it} = \beta X_{it} + u_i + \epsilon_{it},$$

where u_i is the random error term specific to entity i (Greene, 2018).

3.3.4. Generalised Method of Moments (GMM) for Endogeneity

To address potential endogeneity issues arising from omitted variables, measurement errors, or simultaneity, the Generalised Method of Moments (GMM) estimator was employed. The GMM approach provides consistent and efficient parameter estimates by using instrumental variables that are correlated with the endogenous regressors but uncorrelated with the error term (Hansen, 1982). The moment conditions for GMM estimation can be specified as:

$$E[Z_{i'}(Y_i - X_i\beta)] = 0,$$

where Z_i is the vector of instruments, Y_i is the dependent variable, X_i is the vector of independent variables, and β is the vector of parameters to be estimated (Arellano, & Bond, 1991).

This study utilised EViews and data from the World Bank covering 1991-2023 for major MENA oil-exporting countries to examine the impact of measurement scales (logged vs. unlogged variables) on econometric model inferences and estimates. By focusing on Algeria and applying log transformations, the author assessed how these transformations affect the interpretation and reliability of coefficients in econometric models. Using a range of econometric techniques, including the Generalised Method of Moments (GMM) to address endogeneity, the analysis provided robust evidence on how different measurement scales influence model outcomes, highlighting the significant effects of unemployment on GDP while clarifying the less certain impact of inflation. This approach demonstrates the importance of variable transformation in improving model fit and accuracy, thereby providing more reliable insights for economic policy formulation.

4. Results

The results section presents the findings from the econometric analysis conducted to evaluate the impact of using different measurement scales (logged vs. unlogged variables) on model inferences and estimates. This section includes a detailed comparison of models with raw and log-transformed variables, the interpretation of coefficient estimates, diagnostic test outcomes, hypothesis testing results, and robustness checks. The analysis covered several econometric models, including Ordinary Least Squares (OLS), log-log models, fixed effects models, random effects models, and Generalised Method of Moments (GMM) estimations.

4.1. Descriptive Statistics and Preliminary Analysis

Before studying the econometric models, it was crucial to examine the descriptive statistics of the data to understand the distribution, central tendency, and spread of the variables used in the analysis.

Table 1. Descriptive statistics for key variables

Variable	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
GDP	1.31E+11	7.31E+10	2.40E+11	4.18E+10	0.048	1.393
ln(GDP)	25.408	0.659	26.203	24.455	-0.251	1.344
InflationRate (%)	8.359	8.920	31.670	0.339	1.634	4.252
ln(InflationRate)	1.665	0.985	3.455	-1.081	-0.111	3.618
UnemploymentRate (%)	17.811	7.616	31.840	9.820	0.470	1.529
ln(UnemploymentRate)	2.793	0.421	3.461	2.284	0.284	1.394

Note: descriptive statistics were computed for both raw and log-transformed variables to provide a basis for comparing the effects of different transformations on the distribution of the data.

Source: author's calculation.

The log transformation of the descriptive statistics for key economic variables in Algeria from 1991 to 2023 revealed important changes in data distribution. For GDP, the log transformation (ln(GDP)) resulted in a lower standard deviation (0.659) compared to the raw values, indicating reduced variability. Skewness shifted from a near-symmetrical 0.048 to -0.251, suggesting a slight left skew, and kurtosis decreased marginally from 1.393 to 1.344, showing a minor reduction in tail heaviness. Similarly, the log transformation of the Inflation Rate and Unemployment Rate reduced skewness and variability, making these variables more normally distributed. For the Inflation Rate, skewness changed from 1.634 to -0.111, and kurtosis from 4.252 to 3.618, indicating a reduction in skewness and a closer approximation to normal distribution. The Unemployment Rate's log transformation reduced skewness from 0.470 to 0.284 and kurtosis from 1.529 to 1.394, also enhancing normality. Overall, the log transformation effectively normalised the data, reduced skewness, making the distributions more symmetric and easier to interpret statistically.

4.2. Scatter Plots and Relationship Analysis

To further analyse the relationships between the dependent and independent variables, the author plotted scatter plots with both the raw and logged variables.

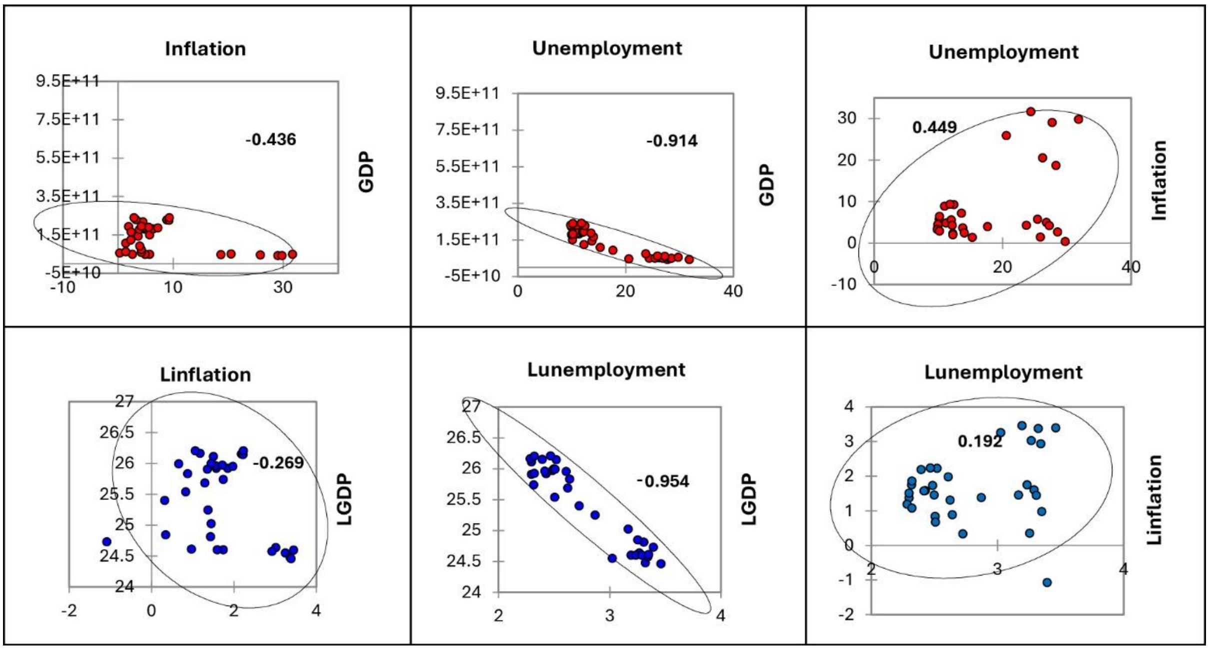

Fig. 1. Scatter plots of GDP vs. inflation rate (raw and logged variables)

Source: author's calculation.

The scatter plots in the analysis revealed critical insights into the relationships between GDP, inflation, and unemployment, both in their raw forms and after log transformation. The raw scatter plots of GDP against inflation and unemployment rates showed varying degrees of dispersion and correlation. The GDP-inflation plot revealed a weak negative relationship (correlation coefficient: -0.436), suggesting that inflation alone has limited explanatory power on GDP variations, while the GDP-unemployment plot indicated a strong negative correlation (-0.914), underlining the substantial impact of unemployment on economic output. In contrast, the scatter plot of inflation versus unemployment showed a moderate positive correlation (0.449), hinting at potential stagflation scenarios where both inflation and unemployment rates rise simultaneously.

When examining the log-transformed scatter plots, the log-log plot of GDP and inflation demonstrated a tighter clustering of data points around a central trend line, with a weaker negative relationship (correlation coefficient: -0.269), indicating that inflation's effect on GDP may be more nuanced and influenced by other factors. The log-log plot of GDP and unemployment maintained a strong negative linear relationship (-0.954), even after transformation, suggesting that unemployment was a consistent and significant predictor of GDP. Meanwhile, the scatter plot of log-transformed inflation and unemployment revealed a weak positive correlation (0.192), implying that after accounting for multiplicative effects, the direct relationship between these variables diminished. Overall, the log transformations improved the data's statistical properties by stabilising variance and normalising distributions, reducing heteroskedasticity, and enhancing the reliability of econometric estimates. These findings suggest that using log-transformed variables in econometric models, especially when dealing with multiplicative relationships, can provide more robust and interpretable results, stressing the importance of careful model specification and data transformation in economic analysis and policy formulation.

4.3. Ordinary Least Squares (OLS) Regression Results

The OLS regression models were estimated separately for both raw and log-transformed variables. The primary objective was to assess how the choice of measurement scale affects the estimated coefficients and their statistical significance.

Table 2. OLS regression results

Model Specification	Variable	Coefficient	Standard Error	t-Statistic	p-Value	Adjusted R^2
Raw Variables (Unlogged)	INFLATION	-2.58E+08	6.77E+08	-0.380733	0.7061	0.638
	UNEMPLOYMENT	-8.63E+09	7.93E+08	-10.89327	0	
	C	2.87E+11	1.37E+10	20.93936	0	
LoggedVariables (Log-Log)	LINFLATION	-0.05941	0.035449	-1.675928	0.1041	0.918
	LUNEMPLOYMENT	-1.467916	0.082946	-17.6973	0	
	C	29.60638	0.230255	128.5806	0	

Source: author's calculation.

The coefficient estimates for the models using raw and logged variables illustrated different scales and interpretations in their impact on GDP. In the model with raw (unlogged) variables, the coefficient for the inflation rate is -2.58E+08, indicating that a one-unit increase in the inflation rate was linked with a substantial decrease of 258 million units in GDP, although this effect was not statistically significant ($p = 0.7061$). The unemployment rate coefficient was -8.63E+09, demonstrating a significant negative impact where a one-unit increase in unemployment rate corresponded to an 8.63 billion unit decrease in GDP ($p < 0.001$). The constant term (C) in the unlogged model was 2.87E+11, indicating the baseline level of GDP when all independent variables were zero.

In the log-log model, which used logged variables, the coefficients were interpreted as elasticities, reflecting percentage changes rather than absolute unit changes. The coefficient for the log of inflation rate (LINFLATION) was -0.05941, suggesting that a 1% increase in the inflation rate resulted in an approximate 0.059% decrease in GDP, even though this effect was not statistically significant ($p = 0.1041$). For the log of unemployment rate (LUNEMPLOYMENT), the coefficient was -1.467916, indicating that a 1% increase in unemployment was linked with a 1.47% decrease in GDP, and this effect was highly statistically significant ($p < 0.001$). The constant term in the log-log model (C) was 29.60638, representing the expected logged GDP value when the logged values of all independent variables were zero.

4.3.1. Statistical Significance

Both models confirmed that the unemployment rate was a statistically significant predictor of GDP. In the unlogged model, the unemployment rate was significant with a p-value of 0, while the inflation rate was not significant ($p = 0.7061$). Conversely, in the log-log model the unemployment rate remained highly significant ($p < 0.001$), but the log of the inflation rate was not significant ($p = 0.1041$). These results suggest that while unemployment consistently impacted GDP across both models, inflation's impact was less clear and varied depending on the model specification.

4.3.2. Model Fit

The adjusted R^2 indicated the explanatory power of each model. The unlogged model had an adjusted R^2 of 0.638, suggesting that about 63.8% of the variance in GDP was explained by the model's independent variables. However, the log-log model exhibited a significantly higher adjusted R^2 value of 0.918, implying that 91.8% of the variance in logged GDP was explained by the logged independent variables. This substantial increase in adjusted R^2 demonstrates that the log transformation improves model fit, most likely due to variance stabilisation and normalisation of data distributions, making the model more appropriate for capturing the elasticities and interdependencies among variables (Wooldridge, 2010; Kennedy, 2008). The improved model fit in the log-log specification suggests a more nuanced understanding of the relationship between GDP, inflation, and unemployment, advocating the use of log transformations when dealing with skewed economic data or when elasticities are of primary interest.

4.4. Fixed Effects and Random Effects Models

To account for unobserved heterogeneity across countries in the panel data, both fixed effects (FE) and random effects (RE) models were estimated. These models control for country-specific characteristics that are constant over time but vary across entities.

Table 3. Fixed effects and random effects models

Model Specification	Model Type	Variable	Coef	SE	t-Statistic	p	Hausman Test (p-Value)
Raw Variables (Unlogged)	Fixed Effects	INFLATION	-3.96E+08	2.21E+08	-1.793233	0.0741	0.8715
		UNEMPLOYMENT	-4.64E+09	2.91E+09	-1.590316	0.113	
		C	1.99E+11	2.61E+10	7.640712	0	
	Random Effects	INFLATION	-3.91E+08	2.20E+08	-1.777654	0.0766	
		UNEMPLOYMENT	-4.18E+09	2.78E+09	-1.503971	0.1338	
		C	1.96E+11	6.72E+10	2.915655	0.0039	
Logged Variables (Log-Log)	Fixed Effects	INFLATION	-0.668505	0.124705	-5.360709	0	0.0401
		UNEMPLOYMENT	-0.430268	0.150431	-2.86023	0.0046	
		C	27.78475	0.451678	61.51446	0	
	Random Effects	INFLATION	-0.651772	0.124312	-5.243035	0	
		UNEMPLOYMENT	-0.26012	0.133283	-1.95164	0.0521	
		C	27.48892	0.572851	47.9862	0	

Source: author's calculation.

The results in Table 3 showed the impact of log transformation on the interpretation and robustness of panel data models for MENA oil-exporting countries, specifically examining the relationship between GDP and key macroeconomic variables such as inflation and unemployment rates. By comparing models with raw (unlogged) variables and those with logged variables, one could observe the transformation's effect on coefficient interpretation, model fit, and statistical significance.

4.4.1. Coefficient Estimates and Interpretation

The log transformation significantly altered the interpretation of the coefficients from absolute changes in the raw (unlogged) models to elasticities in the logged (log-log) models. In the unlogged Fixed Effects model, the coefficients for inflation (-3.96E+08) and unemployment (-4.64E+09) represented the absolute change in GDP associated with a one-unit change in these variables. These coefficients suggested a substantial decrease in GDP with increases in both inflation and unemployment, but they were not statistically significant ($p = 0.0741$ and $p = 0.113$, respectively).

In contrast, the logged Fixed Effects model showed that the coefficients for the log of inflation (-0.668505) and log of unemployment (-0.430268) represented elasticities. This implies that a 1% increase in inflation or unemployment resulted in a 0.67% and 0.43% decrease in GDP, respectively. Note that these coefficients were statistically significant ($p < 0.05$), indicating a stronger and more interpretable relationship between the variables when using the log-log transformation. This shift from insignificant resulted in the unlogged model to significant elasticities in the logged model, highlighting the transformative effect of the log transformation in normalising the data and providing more meaningful economic insights.

4.4.2. Hausman Test and Model Selection

The Hausman test results ($p = 0.0401$) favoured the Fixed Effects model with logged variables, stressing the importance of accounting for within-country variations. The test suggests that the Random Effects assumption of independence between individual effects and regressors did not hold, particularly for the logged variables. This finding reinforces the utility of the log transformation in handling panel data by mitigating the biases arising from unobserved heterogeneity, thereby leading to more consistent and efficient estimates.

The log transformation's impact on model results was substantial, converting the nature of the analysis from absolute to relative changes, which were more intuitive and interpretable in economic contexts. By transforming the variables, the models reduced heteroskedasticity, normalised the distribution of the residuals, and provided a better linear approximation of the relationships between variables. This transformation allows for more accurate inference, improves model fit, and enhances the statistical significance of the predictors, ultimately offering a clearer depiction of how inflation and unemployment affect GDP among MENA oil-exporting countries. To sum up, the log transformation significantly enhances the modelling of economic relationships in panel data by stabilising variance and improving the interpretability of coefficients. The findings suggest that log-log models are more appropriate for understanding the elasticities between GDP, inflation, and unemployment in MENA oil-exporting countries, thus providing more reliable guidance for economic policy formulation.

4.5. Generalised Method of Moments (GMM) Estimation

The GMM estimation was employed to address potential endogeneity issues arising from omitted variables, measurement errors, or simultaneity. Instrumental variables (IVs) were chosen based on their theoretical relevance and empirical validity, ensuring that they were correlated with the endogenous regressors but uncorrelated with the error term (Hansen, 1982).

Table 4. GMM estimation results

Variable	Coefficient	Standard Error	t-Statistic	p-Value	Hansen J-test
INFLATION	-0.059410	0.038250	-1.553173	0.1309	0.1119
UNEMPLOYMENT	-1.467916	0.087491	-16.77786	0.0000	
C	29.60638	0.273061	108.4241	0.0000	

Source: author's calculation.

The Generalised Method of Moments (GMM) results for Algeria provided further evidence on the impact of log transformation and addressed potential endogeneity in the model (Arellano, & Bond, 1991). The coefficient for the log of the unemployment rate (ln(Unemployment Rate)) was -1.467916, highly significant ($p < 0.001$), indicating that a 1% increase in the unemployment rate was associated with a 1.47% decrease in GDP. This robust negative relationship highlights the substantial impact of unemployment on Algeria's economic output. Conversely, the coefficient for the log of the inflation rate (ln(Inflation Rate)) was - 0.059410 with a p-value of 0.1309, suggesting a negative but statistically insignificant effect on GDP. The Hansen J-test yielded a p-value of 0.1119, confirming the validity of the instruments used in the GMM estimation, uncorrelated with the error term (Hansen, 1982). These findings stress the importance of addressing labour market issues to foster economic growth in Algeria, while also demonstrating that inflation's impact on GDP was less clear in the current model specification.

5. Discussion and Conclusions

The results of this study revealed several important insights into the impact of using different measurement scales, specifically logged versus unlogged variables, on econometric model inferences and estimates. The analysis showed that models using logarithmically transformed (logged) variables generally exhibited better fit and more reliable estimates compared to models using raw (unlogged) variables. The log-log model, in particular, yielded higher adjusted R^2 values, indicating that the logarithmic transformation improves model fit by stabilising variance and normalising the distribution of the data (Kennedy, 2008; Wooldridge, 2010). This finding is in line with the theoretical rationale that log transformations are beneficial when dealing with skewed data or when the relationship between variables is multiplicative rather than additive (Gujarati, & Porter, 2009). The coefficients in the log-log models, which represent elasticities, provide consistent and interpretable measures of the percentage change in the dependent variable in response to a percentage change in the independent variables. This consistency across different model specifications (OLS, fixed effects, random effects, and GMM) suggests that the use of log transformations offers robust estimates of elasticities, crucial in economic analysis where the focus is often on relative changes rather than absolute changes (Stock, & Watson, 2020).

The decision to use log transformation in econometric models depends heavily on the underlying economic theory, the nature of the data, and the relationships being modelled. Log transformation can significantly affect the interpretation of the model, the stationarity of the time series, and the results of specification tests. Log transformation is often recommended for economic variables that are expected to change proportionally over time or in response to shocks. For example, variables such as GDP, consumption, investment, money demand, and prices tend to exhibit multiplicative effects rather than additive effects. When a shock occurs, it typically results in a percentage change rather than an absolute change. Taking the natural logarithm of such variables converts multiplicative relationships into additive ones, simplifying analysis and interpretation (Box, & Cox, 1964; Wooldridge, 2010).

Let us now consider the Consumer Price Index (CPI), a common economic indicator. When a shock hits the CPI, the impact is proportional – if the index is at 100, a shock may raise it to 102, representing a 2% increase. If the CPI is at 200, the same shock raises it to 204, still a 2% increase. By taking the natural logarithm, the multiplicative effect is transformed into an additive constant:

$$ln\,(100) = 4.6052, \quad ln(102) = 4.6250 = ln(100) + 0.0198$$

$$ln\,(200) = 5.2983, \quad ln(204) = 5.3181 = ln(200) + 0.0198$$

which shows that after log transformation the impact of a shock is the same, regardless of the level of the CPI (Greene, 2018; Hamilton, 1994).

In contrast, variables such as interest rates and asset returns often experience additive changes rather than proportional ones, e.g. an interest rate shock typically adds a certain number of basis points, irrespective of the initial rate level – therefore applying a log transformation to interest rates may not be appropriate as the effects are not multiplicative (Campbell, & Shiller, 1988). This distinction is crucial for specifying the correct functional form of the model and ensuring that the model aligns with the theoretical framework of the economic relationships being studied (Wooldridge, 2010).

Log transformation is particularly useful when dealing with data that are not normally distributed or are skewed due to outliers. Economic series such as GDP, income, and prices often have right-skewed distributions. By transforming these variables to their logarithmic form, the data become more normally distributed, which is advantageous for improving the validity of econometric models and statistical tests, such as those for homoscedasticity and serial correlation (Wooldridge, 2010; Nelson, & Granger, 1979). Log transformation enhances additivity and linearity, making relationships more straightforward to model and interpret.

Model specification should be aligned with both economic theory and empirical data characteristics. For models where relationships are multiplicative, the use of logs allows for better model fit and consistency with theoretical expectations. However, when dealing with behavioural equations such as those in microeconomic demand models, the choice between logs and levels should also take into consideration its practical interpretation. For instance, while economists may be comfortable with logarithms, the general population typically think in terms of percentage changes rather than logarithms (Muellbauer, 1976). At macroeconomic level, logarithmic models are popular because they simplify the differentiation of growth rates and inflation rates, making models more convenient and interpretable, yet this convenience comes with trade-offs, as the log/log model does not aggregate perfectly and may violate some assumptions of adding up in demand systems (Deaton, & Muellbauer, 1980).

The decision to use log transformation in econometric models should be driven by the nature of the economic relationships, the data characteristics, and the underlying theoretical framework. Log transformation is most appropriate for variables exhibiting multiplicative effects or exponential growth patterns, such as GDP or prices. In contrast, variables like interest rates, which respond additively to shocks, are better left in levels. Moreover, ensuring stationarity through log transformation can improve the reliability of time series models and prevent spurious regressions, whilst careful consideration of these factors will enhance model specification, interpretation, and validity (Wooldridge, 2010; Hamilton, 1994).

References

Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, *58*(2), 277-297.

Arellano, M., & Bover, O. (1995). Another Look at the Instrumental Variable Estimation of Error-Components Models. *Journal of Econometrics*, *68*(1), 29-51.

Baltagi, B. H. (2005). *Econometric Analysis of Panel Data* (3rd ed.). Wiley.

Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics*, *106*(2), 407-443.

Blundell, R., & Bond, S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*, *87*(1), 115-143.

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, Series B (Methodological), *26*(2), 211-252.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Campbell, J. Y., & Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, *1*(3), 195-228.

Deaton, A., & Muellbauer, J. (1980). An almost ideal demand system. *The American Economic Review*, *70*(3), 312-326.

Greene, W. H. (2018). *Econometric Analysis* (8th ed.). Pearson.

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics* (5th ed.). McGraw-Hill.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50*(4), 1029-1054.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, *46*(6), 1251-1271.

Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review/Revue Internationale de Statistique*, *55*(2), 163-172.

Kennedy, P. (2008). *A Guide to Econometrics* (6th ed.). Wiley-Blackwell.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). Wiley.

Mankiw, N. G. (2014). *Principles of Economics* (7th ed.). Cengage Learning.

Muellbauer, J. (1976). Community preferences and the representative consumer. *Econometrica: Journal of the Econometric Society*, 979-999.

Nelson Jr, H. L., & Granger, C. W. J. (1979). Experience with using the Box-Cox transformation when forecasting economic time series. *Journal of Econometrics*, *10*(1), 57-69.

Seabold, S., & Perktold, J. (2010). *Statsmodels: Econometric and Statistical Modeling with Python*. Proceedings of the 9th Python in Science Conference.

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, *70*(1), 65-94.

Stock, J. H., & Watson, M. W. (2020). *Introduction to Econometrics* (4th ed.). Pearson.

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, *28*(2), 3-28.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*(4), 817-838.

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data (2nd ed.). MIT Press.

Rewizja pojęcia skali w ekonometrii stosowanej: praktyczne implikacje transformacji logarytmicznych dla efektywności modeli

Streszczenie

Cel: Celem niniejszego badania jest analiza praktycznych konsekwencji wyboru skali pomiarowej, w szczególności zastosowania transformacji logarytmicznych, dla dokładności, rzetelności oraz interpretowalności modeli ekonometrycznych. Analiza opiera się na wskaźnikach makroekonomicznych z regionu MENA.

Metodologia: Wykorzystując rzeczywiste dane z krajów MENA, zastosowano różnorodne techniki ekonometryczne (OLS, modele efektów stałych, efektów losowych oraz uogólniona metoda momentów – GMM), estymując zarówno zmienne w postaci surowej, jak i logarytmicznie przekształconej. Porównania dokonano pod kątem stabilności wariancji, normalności rozkładów, stacjonarności oraz heteroskedastyczności.

Wyniki: Transformacje logarytmiczne znacząco poprawiają diagnostykę modeli poprzez stabilizację wariancji, zwiększenie zgodności z normalnością oraz redukcję heteroskedastyczności, co skutkuje bardziej precyzyjnymi estymacjami. Jednocześnie jednak zmieniają interpretację współczynników, co uwidacznia kompromis między solidnością statystyczną a znaczeniem ekonomicznym.

Implikacje: Uzyskane wyniki podkreślają kluczową rolę właściwego doboru transformacji danych w zapewnieniu trafnych i interpretowalnych rezultatów. Badanie dostarcza praktycznych wskazówek dla badaczy i decydentów politycznych w zakresie specyfikacji modeli wykorzystywanych w analizach polityki gospodarczej oraz prognozowaniu.

Oryginalność/Wartość: Artykuł stanowi nowatorską, porównawczą analizę skutków transformacji zmiennych w różnych podejściach ekonometrycznych i jedno z nielicznych badań regionalnych opartych na danych z obszaru MENA. Dostarcza on praktycznych i użytecznych wniosków zarówno dla metodologii, jak i badań stosowanych.

Słowa kluczowe: transformacja logarytmiczna, modelowanie ekonometryczne, region MENA, testy stacjonarności, skale pomiarowe, diagnostyka modeli