Reassessing the Impact of Labor and Innovation on SEZ Output: An Econometric Approach.

Murodil Isroiljonov Qosimjon o'g'li
Undergraduate Student, Faculty of Economics
Tashkent State University of Economics
ORCID: https://orcid.org/0009-0009-5149-1341
Email: murodil.isroiljonov06@gmail.com
Phone number: +998 (90) 770-72-01

Abstract: This study reassesses the determinants of production output in Special Economic Zones (SEZs), focusing on the roles of labor and innovation. Using annual data from 2010 to 2025 of Qashqadarya, the analysis applies Ordinary Least Squares (OLS) regression, including both linear and log-linear specifications. Initial results indicate a high explanatory power, with R^2 exceeding 0.96, suggesting strong relationships between variables. However, further diagnostic testing reveals severe multicollinearity among independent variables and the presence of strong time trends. After refining the model and applying a log-linear specification, labor remains a statistically significant determinant of output, while innovation loses significance when controlling for time effects. This suggests that the apparent influence of innovation may be driven by shared upward trends rather than a direct causal relationship. The findings highlight the importance of rigorous diagnostic testing in econometric modeling, demonstrating that high R^2 values can be misleading. The study concludes that careful model specification is essential to avoid incorrect economic interpretations.

Key words: Special Economic Zones (SEZs), multicollinearity, Log-Linear Specification, labor, Ordinary Least Squares (OLS), correlation, innovation.

1 INTRODUCTION.

Special Economic Zones (SEZs) are widely recognized as engines of economic growth, contributing to industrial development, export expansion, and employment generation. Understanding the key determinants of production output within SEZs is essential for designing effective economic policies.

Traditional economic theory suggests that production output is driven by factors such as labor, investment, innovation, and exports. Empirical studies often report strong positive relationships between these variables and economic growth. However, many of these studies overlook important econometric issues, such as multicollinearity and time trends, which can distort results and lead to misleading conclusions.

A key limitation in existing research is the assumption that high statistical significance implies a true causal relationship. In reality, variables that grow over time may appear strongly related even when no direct causal link exists. This study aims to reassess the impact of labor and innovation on SEZ production output using a rigorous econometric framework.

Research Question: What are the true effects of labor and innovation on SEZ production output?

Hypothesis: Both labor and innovation are expected to have positive effects on output.

Thesis Statement: Although initial regression results suggest strong positive relationships, further analysis reveals that these relationships are partly driven by multicollinearity and time trends, requiring cautious interpretation.

2 LITERATURE REVIEW.

Economic theory emphasizes the importance of factor inputs in determining production output. Labor contributes directly to production capacity, while innovation enhances productivity through technological improvements. Endogenous growth theory highlights innovation as a key driver of long-term economic growth.

Empirical studies generally find positive relationships between labor, innovation, and output. However, more recent research has identified significant methodological challenges. In particular, multicollinearity among economic variables often leads to unstable coefficient estimates, making it difficult to isolate individual effects. Additionally, time-series data frequently exhibit strong trends, which can result in spurious regression outcomes.

Despite these concerns, many studies continue to rely on high R^2 values as evidence of strong relationships, without adequately testing model assumptions. This study addresses this gap by applying comprehensive diagnostic testing to evaluate the robustness of regression results.

3 METHODOLOGY

3.1 Research Design

This study employs a quantitative econometric approach using Ordinary Least Squares (OLS) regression to examine the relationship between production output and its determinants in SEZs. Both linear and log-linear specifications are estimated to test robustness and improve interpretability of coefficients.

3.2 Data

The analysis uses annual data from 2010 to 2025, consisting of 16 observations. The variables include:

Y – Production output

X1 – Labor

X2 – Innovation

X3 – Investment

X4 – Exports

3.3 Model Specification

Two main models are estimated:

Linear model:

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \beta_4 X4 + \varepsilon$$

To improve interpretability and address potential non-linearity, a log-linear specification is employed. The preferred model is defined as:

$$lnY = \beta_0 + \beta_1\, lnX1 + \beta_2\, lnX2 + \varepsilon$$

3.4 Diagnostic tests

The following tests are conducted:

(1) Multicollinearity: Variance Inflation Factor (VIF) used to detect whether independent variables contain overlapping information. High VIF (>10) implies variables move together, this makes coefficient estimates unstable and unreliable.

(2) Heteroskedasticity: Breusch–Pagan and White tests used to check whether error variance is constant. If violated → standard errors are biased, this affects hypothesis testing (t-tests, p-values).

(3) Autocorrelation: Durbin–Watson and Breusch–Godfrey tests used in time-series context to test whether residuals are correlated across time. If present OLS standard errors become invalid, especially important in annual macroeconomic data.

(4) Normality tests: Shapiro–Wilk and Skewness/Kurtosis tests used to assess whether residuals follow normal distribution. Important for validity of inference in small samples. Weak test in large samples, but relevant here due to n=16.

(5) Specification tests: Ramsey RESET and Linktest used to detect whether the model is misspecified or missing nonlinear relationships.

3.5 Time Trend Inclusion

A deterministic time variable (t) is included to control for common growth patterns across variables because it might reduce spurious correlation and separate structural relationships from time-driven movement.

4 RESULTS

4.1 Descriptive statistics

The descriptive statistics indicate that all variables exhibit substantial variation and an upward trend over time, that we can see in the following table:

Variable	Obs	Mean	Std. dev.	Min	Max
Y	16	314.8375	156.9362	45.6	528.2
X1	16	1812.625	692.4616	435	2737
X2	16	1605.406	451.8246	654	2234.8
X3	16	935.1812	393.5763	214.8	1432.8
X4	16	20.58438	14.62684	3.32	40.3

Table 1. Descriptive statistics

Overall, variables show strong upward variation over time. Output ranges from 45.6 to 528.2, while input variables also exhibit increasing trends. This indicates that all variables share a common growth structure, suggesting potential time dependence in the data.

4.2 Correlation results

The correlation matrix and scatterplot matrix indicate very strong positive relationships between the dependent variable and all explanatory variables, as we can see in the Table 2:

	Y	X1	X2	X3	X4
Y	1.0000				
X1	0.9730*	1.0000			
	0.0000				
X2	0.9708*	0.9695*	1.0000		
	0.0000	0.0000			
X3	0.9627*	0.9663*	0.9874*	1.0000	
	0.0000	0.0000	0.0000		
X4	0.9154*	0.8805*	0.9197*	0.9332*	1.0000
	0.0000	0.0000	0.0000	0.0000	

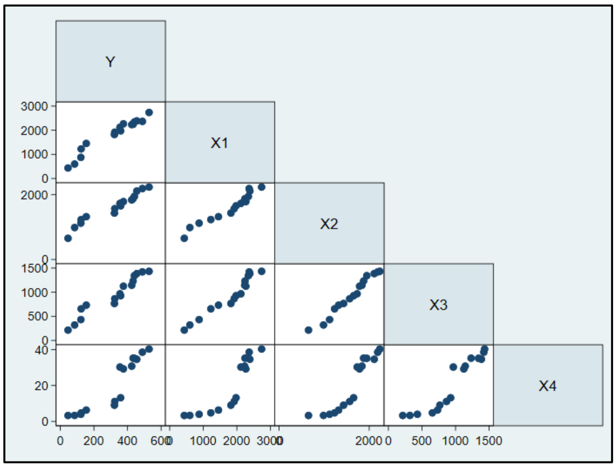

Table 2. Correlation matrix and scatterplot matrix

The correlations among independent variables are also extremely high, particularly between innovation (X2) and investment (X3) which equal to 0.9874.

Correlations between X's:

X2 & X3 = 0.9874

X1 & X2 = 0.9695

X1 & X3 = 0.9663

X3 & X4 = 0.9332

These values indicate strong co-movement among variables, suggesting potential multicollinearity and shared trend dynamics rather than independent relationships.

4.3 Baseline regression

The baseline OLS regression yields a high coefficient of determination (R^2 = 0.9662), and the model is jointly significant (F = 78.71, $p < 0.001$). However, individual coefficient estimates are not uniformly significant.

Labor (X1) enters positively and is statistically significant (β = 0.140, p = 0.025), whereas innovation (X2), investment (X3), and exports (X4) are not tatistically significant at conventional levels. Notably, the coefficient on investment is negative, which is inconsistent with standard production theory.

$$Y = -129.146 + 0.140X1 + 0.154X2 + \beta_3 X3 - 0.119X4 + \varepsilon$$

Source	SS	df	MS			
Model	356962.549	4	89240.6374	Number of obs	=	16
Residual	12472.188	11	1133.83528	F(4, 11)	=	78.71
				Prob > F	=	0.0000
Total	369434.737	15	24628.9825	R-squared	=	0.9662
				Adj R-squared	=	0.9540
				Root MSE	=	33.672

Y	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
X1	.1401829	.054231	2.58	0.025	.0208212	.2595445
X2	.1545105	.1316508	1.17	0.265	-.1352509	.4442719
X3	-.1188116	.1604609	-0.74	0.475	-.4719836	.2343604
X4	2.571984	1.703473	1.51	0.159	-1.177335	6.321303
_cons	-129.1459	74.8849	-1.72	0.113	-293.9665	35.67462

Table 3. Regression Model

These results indicate that, despite strong overall model fit, the contribution of individual regressors is not precisely identified.

4.4 Multicollinearity Diagnostics

Variance Inflation Factors (VIF) reported in Table 4 confirm the presence of severe multicollinearity. The VIF values for investment (52.76) and innovation (46.81) substantially exceed conventional thresholds, indicating that explanatory variables are highly interdependent.

Variable	VIF	1/VIF
X3	52.76	0.018952
X2	46.81	0.021363
X1	18.66	0.053601
X4	8.21	0.121755
Mean VIF	31.61	

Table 4. Variance Inflation Factor (VIF)

Under such conditions, coefficient estimates become unstable, and standard errors are inflated, undermining the reliability of statistical inference.

4.5 Full Log-Linear Model

The full log-linear specification including all explanatory variables was estimated as follows:

$$lnY = -8.932 + 1.378\ lnX1 + 1.732\ lnX2 - 1.317\ lnX3 + 0.169\ lnX4 + \varepsilon$$

Source	SS	df	MS			
				Number of obs	=	16
				F(4, 11)	=	162.65
Model	7.96232198	4	1.9905805	Prob > F	=	0.0000
Residual	.134624867	11	.012238624	R-squared	=	0.9834
				Adj R-squared	=	0.9773
Total	8.09694685	15	.539796456	Root MSE	=	.11063

lnY	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnX1	1.378209	.311784	4.42	0.001	.6919771	2.064441
lnX2	1.731775	.5403637	3.20	0.008	.5424425	2.921107
lnX3	-1.317381	.4591442	-2.87	0.015	-2.327951	-.3068117
lnX4	.169165	.077048	2.20	0.050	-.0004164	.3387465
_cons	-8.93194	2.120609	-4.21	0.001	-13.59937	-4.264512

Table 5. Log-Linear model

The model (Table 5) exhibits a high level of explanatory power ($R^2 = 0.9834$), and several coefficients appear statistically significant. In particular, labor and innovation show positive and significant elasticities, while investment enters with a negative coefficient, which is inconsistent with economic theory.

Despite the strong statistical fit, the reliability of this specification is undermined by severe multicollinearity (Table 6) among the explanatory variables.

This is reflected in extremely high VIF values (mean VIF = 40.41), indicating that the regressors share substantial overlapping information.

Variable	VIF	1/VIF
lnX3	81.02	0.012342
lnX2	39.11	0.025571
lnX1	34.66	0.028851
lnX4	6.85	0.145989
Mean VIF	40.41	

Table 6. Variance Inflation Factor (VIF) after Log-Linear model

As a result, the estimated coefficients are unstable and sensitive to model specification. The negative coefficient on investment, in particular, suggests that the model fails to provide a meaningful economic decomposition of output.

Consequently, the full log-linear specification is not retained as the preferred model, and a reduced specification is considered.

4.6 Log-Linear Specification

Based on diagnostic testing and model refinement, the preferred specification is the log-linear model including labor and innovation. The estimated regression equation is:

$$lnY = -7.262 + 0.729 lnX1 + 1.012\ lnX2 + \varepsilon$$

This specification allows coefficients to be interpreted as elasticities, providing a direct measure of the percentage response of output to changes in inputs.

The estimated elasticity of output with respect to labor is 0.730. This implies that a 1% increase in labor is associated with a 0.73% increase in output, holding innovation constant. This magnitude indicates diminishing returns to labor, which is consistent with standard production theory, where additional labor contributes positively to output but at a decreasing rate.

The elasticity of output with respect to innovation is 1.012, suggesting that a 1% increase in innovation is associated with approximately a 1.01% increase in output. In the baseline specification, this indicates a near-proportional relationship and implies that innovation may play a strong productivity-enhancing role in SEZ output generation.

However, this interpretation should not be taken as definitive evidence of a stable structural relationship. As shown in robustness analysis, the statistical significance and magnitude of the innovation coefficient are sensitive to model specification.

Source	SS	df	MS			
				Number of obs	=	16
				F(2, 13)	=	191.48
Model	7.8311048	2	3.9155524	Prob > F	=	0.0000
Residual	.265842045	13	.020449388	R-squared	=	0.9672
				Adj R-squared	=	0.9621
Total	8.09694685	15	.539796456	Root MSE	=	.143

lnY	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnX1	.7296629	.2764835	2.64	0.020	.1323566	1.326969
lnX2	1.012335	.4511211	2.24	0.043	.0377472	1.986923
_cons	-7.262418	1.421846	-5.11	0.000	-10.33413	-4.190706

Table 7. Log-Linear Specification model

Despite the improved interpretability of the log-linear specification, multicollinearity remains a significant concern. The Variance Inflation Factor (VIF) values for both labor and innovation are approximately 16.31 (Table 8), exceeding the conventional threshold of 10.

Variable	VIF	1/VIF
lnX1	16.31	0.061303
lnX2	16.31	0.061303
Mean VIF	16.31	

Table 8. Variance Inflation Factor (VIF)

This indicates a high degree of linear dependence between explanatory variables, meaning that labor and innovation contain overlapping information. In such conditions, regression coefficients may remain statistically significant while still being unstable and sensitive to small changes in model specification.

Consequently, the independent contribution of each variable cannot be fully isolated, and coefficient interpretation must be treated with caution.

Further robustness analysis introduces a deterministic time trend (t) to account for common growth dynamics across variables. The results (Table 9) indicate that the coefficient on innovation loses statistical significance (β = 0.723, p = 0.313) once the time trend is included in the model, while labor remains relatively stable (β = 0.797, p = 0.024).

Source	SS	df	MS			
				Number of obs	=	16
				F(3, 12)	=	121.14
Model	7.8381316	3	2.61271053	Prob > F	=	0.0000
Residual	.258815251	12	.021567938	R-squared	=	0.9680
				Adj R-squared	=	0.9600
Total	8.09694685	15	.539796456	Root MSE	=	.14686

lnY	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnX1	.7972924	.3076736	2.59	0.024	.1269292	1.467656
lnX2	.7228262	.6869525	1.05	0.313	-.7739148	2.219567
t	.0136017	.0238296	0.57	0.579	-.0383187	.065522
_cons	-5.754459	3.01858	-1.91	0.081	-12.33138	.8224609

Table 9. Log-Linear Specification model after time trend

This suggests that the previously observed effect of innovation may partly reflect shared temporal movement rather than a purely independent structural relationship. In other words, innovation and output appear to co-move over time, and part of their correlation is absorbed by the underlying growth trend rather than representing a direct causal linkage.

5 DISCUSSION

The empirical results provide mixed evidence regarding the determinants of SEZ production output. Across specifications, labor consistently emerges as the most stable and robust explanatory variable. Its coefficient remains positive and statistically significant in all models, including those incorporating log transformation and time trend adjustments. This stability suggests that labor constitutes a fundamental and structurally identifiable input in SEZ production processes.

In contrast, the role of innovation is substantially less stable. Although initial OLS and log-linear specifications indicate a statistically significant and economically strong positive relationship between innovation and output, this relationship weakens once temporal dynamics are explicitly controlled. Specifically, the inclusion of a deterministic time trend eliminates the statistical significance of innovation, implying that its baseline effect is not robust.

This pattern indicates that innovation and output share a strong common time trajectory. In such settings, standard OLS estimation tends to attribute shared growth patterns to individual regressors, thereby overstating their structural importance. Once the common trend component is isolated, the independent explanatory power of innovation declines substantially.

A further methodological concern arises from severe multicollinearity among explanatory variables. Extremely high VIF values indicate that labor, innovation, investment, and exports move closely together over time, reflecting a shared underlying growth process rather than independent variation. This undermines the ability of the model to disentangle marginal effects and results in unstable coefficient estimates across specifications.

Consequently, the high explanatory power observed in all models ($R^2 > 0.96$) should not be interpreted as evidence of strong causal structure. Instead, it largely reflects common trending behavior among variables. This highlights a key econometric limitation in macro-level time series analysis: high fit does not necessarily imply correct specification.

Overall, the findings suggest that while labor represents a structurally meaningful driver of SEZ output, the estimated effects of innovation and other macro variables are sensitive to specification choices and may reflect shared temporal dynamics rather than independent causal mechanisms.

6 CONCLUSION

This study examined the determinants of SEZ output using OLS and log-linear models with robustness checks. The results show that labor is the only consistently significant and stable determinant across all specifications. In contrast, innovation appears significant only in baseline models but becomes statistically insignificant after controlling for time trends, indicating that its effect is largely driven by shared temporal dynamics rather than a stable structural relationship.

Diagnostic tests further reveal severe multicollinearity among explanatory variables, which reduces the reliability of individual coefficient estimates despite high overall model fit. This suggests that the high R^2 values mainly reflect common trending behavior rather than strong causal structure.

Overall, the findings highlight that labor plays the most robust role in SEZ output, while the effects of innovation and other variables are not stable under alternative specifications and should be interpreted with caution.

7. REFERENCES

1. Gujarati, D. N., & Porter, D. C. (2009). Basic Econometrics. McGraw-Hill Education. (Essential for citing OLS assumptions, Multicollinearity, and VIF thresholds).

2. Wooldridge, J. M. (2015). Introductory Econometrics: A Modern Approach. Cengage Learning. (Used to justify the log-linear specification and time-trend analysis).

3. Romer, D. (2011). Advanced Macroeconomics. McGraw-Hill. (For the theoretical background on Endogenous Growth Theory and Innovation).

4. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48(4), 817–838.

5. Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression, II. Biometrika, 38(1–2), 159–177.

6. Wang, J. (2013). The economic impact of Special Economic Zones: Evidence from Chinese municipalities. Journal of Development Economics, 101, 133–147.

7. Stock, J. H., & Watson, M. W. (2019). Introduction to Econometrics (4th ed.). Pearson Education.

8. Farole, T., & Akinci, G. (Eds.). (2011). Special Economic Zones: Progress, Emerging Challenges, and Future Directions. World Bank Publications.

9.Касимова Ф. Т. ТЕНДЕНЦИИ РАЗВИТИЯ ЦИФРОВОЙ ЭКОНОМИКИ В МАЛОМ БИЗНЕСЕ //Scientific Conference on Multidisciplinary Studies. – 2025. – С. 396-401.

10.Kasimova F. Trends and Prospects for the Development of Digital Marketing Activities in Small Business //International Journal of Early Childhood Special Education. – 2022. – Т. 14. – №. 5.

11.Касимова Ф. Т. СРАВНИТЕЛЬНЫЙ АНАЛИЗ МИРОВЫХ ТЕНДЕНЦИЙ В РАЗВИТИИ СИСТЕМЫ МАРКЕТИНГА В СФЕРЕ ПРЕДПРИНИМАТЕЛЬСТВА В УСЛОВИЯХ ГЛОБАЛИЗАЦИИ МИРОВЫХ РЫНКОВ //Молодой исследователь: вызовы и перспективы. – 2018. – С. 245-250.

12.Касимова Ф. Т. ОСОБЕННОСТИ И ОТЛИЧИТЕЛЬНЫЕ СВОЙСТВА МАЛОГО БИЗНЕСА И ЧАСТНОГО ПРЕДПРИНИМАТЕЛЬСТВА //Интернаука. – 2016. – №. 4-2. – С. 17-19.

13.Tulkunovna K. F. ISSUES OF EVALUATING THE LEVEL OF DEVELOPMENT OF MARKETING ACTIVITIES OF SMALL BUSINESS IN THE REPUBLIC OF UZBEKISTAN //ЕВРАЗИЙСКИЙ СОЮЗ УЧЕНЫХ (ЕСУ). – Т. 17.

14.Kasimova F. Modern Tools of Digital Marketing in Small Business and Entrepreneurship //Asian Journal of Technology & Management Research. – 2022. – Т. 12. – №. 1.